OPTIMAL CODES FOR INFORMATION-THEORETICALLY COVERT COMMUNICATION

A Dissertation
Presented to
The Academic Faculty

By

Ishaque Ashar Kadampot

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology

May 2020

OPTIMAL CODES FOR INFORMATION-THEORETICALLY COVERT COMMUNICATION

Approved by:

Dr. Matthieu Bloch, Advisor
Associate Professor, School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. John Barry
Professor, School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Stephen Ralph
Professor, School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Faramarz Fekri
Professor, School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Yao Xie
Associate Professor, School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: January 08, 2020

“And say: My Lord! Increase me in knowledge”

Holy Qur’an 20:114

To my parents, siblings, and Kochu.

ACKNOWLEDGEMENTS

I am deeply indebted to my advisor, Dr. Matthieu Bloch, for guiding me over the last five years and showing a tremendous amount of patience with me. The completion of my dissertation would not have been possible without the support and motivation he provided throughout this journey. I have learned a lot from the way he worked with his students and the hard work he put into everything. I would like to extend my sincere thanks to Dr. John Barry, Dr. Stephen Ralph, Dr. Faramarz Fekri, and Dr. Yao Xie for being part of my defense committee and agreeing to schedule my defense in short notice. I would also like to thank the academic professionals in the ECE academic office for their help.

Many thanks to Keerthi and Mehrdad, with whom I spent most of my time in the lab, for all those technical and non-technical discussions that made my Ph.D. life memorable. Special thanks to Mehrdad for making my life easier by helping me in solving problems with his mathematical insights. I also wish to extend my thanks to Meng-Che, Nathan, Shi-Yuan, Alex, Rémi, and Guillaume for providing a friendly research environment.

I must also thank Rizwan, Ali Murtaza, and Sujith for helping me to acclimatize with Atlanta when I moved here. I am also grateful to Abdullah, Hasan, Aftab, Muzzammil, Vivek, Yassine, Taofiq, Zohaib, Saad, and Mahmoud for being great roommates. I would also like to extend my gratitude to Abdelkareem, Shabab, Jawad, and all others, who made my stay in Atlanta pleasant.

Most importantly, I am especially grateful to my parents, siblings, and wife for their encouragement and support. Finally, I would like to thank Iqbal for always being there as a true friend.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

We consider a problem of coding for covert communication, which involves ensuring reliable communication between two legitimate parties while simultaneously guaranteeing a low probability of detection by an eavesdropper. Specifically, we develop an optimal low-complexity coding scheme that achieves the information-theoretic limits of covert communication over binary-input discrete memoryless channels. We first demonstrate the non-triviality of designing codes for covert communication by showing the impossibility of achieving information-theoretic limits using linear codes without a shared secret key for a regime in which information theory proves the possibility of covert communication without a secret key. We then circumvent this impossibility by introducing non-linearity into the coding scheme through the use of pulse position modulation (PPM) and multilevel coding (MLC). This MLC-PPM scheme exhibits several appealing properties; in particular, for an appropriate decoder, the channel at a given level is independent of the total number of levels and the codeword length. We exploit these properties to show how one can use families of channel capacity- and channel resolvability-achieving codes to concretely instantiate a covert communication scheme. Further, we extend the MLC-PPM scheme using bi-orthogonal PPM symbols to achieve information-theoretic limits of covert communication over additive white Gaussian channels. Finally, we illustrate the application of this scheme for the secret-key generation problem with a covertness constraint.

CHAPTER 1
INTRODUCTION

The history of secure communication is that of a perpetual contest between codemakers and codebreakers, in which codemakers develop coding schemes with increasing complexity only to have them broken by codebreakers with unexpected consequences. For instance, during the First World War, the British cryptanalysts were successful in breaking the secret diplomatic communication from the German foreign office, popularly known as the Zimmermann telegram, that proposed a military alliance between Germany and Mexico, encouraging direct aggression against the United States. As a result of this, Woodrow Wilson, then president of the United States, decided to join the Allied forces in the war against Germany, which eventually impacted the outcome of the war. Although the encryption schemes that we use today are much more secure than the ones used by the German foreign office, their security is based on the computational complexity of inverting certain mathematical operations. The security of those schemes heavily depends on proving that the computational time required to solve the inverse problem using state-of-the-art computing resources is very high that it is virtually impossible. However, there already exist efficient quantum algorithms that can solve these computationally hard problems much faster. Consequently, the emergence of quantum computers might result in a scenario similar to the breaking of the Zimmermann telegraph. This motivates us to investigate information-theoretically secure communication schemes, which ensure the security of a coding scheme without assuming any limitation on the computational capability of the adversary. Recently, there have been many advances in this area. We will briefly review some of these advances in the next chapter.

Although secure communication protects the message from being decrypted by the adversary, it does not protect users from revealing to the adversary that they are communicat-

ing. In scenarios such as covert military operations, even the presence of communication arouses suspicion to a watchful adversary. In such situations, encryption is not sufficient and it is important to use techniques with low probability of detection (LPD) to exchange strategic information. This type of communication with LPD is also known as covert communication or deniable communication and has attracted renewed interest in information theory. Although there exist many studies that address the information-theoretic limits of various covert communication scenarios, only a handful of studies have addressed the construction of practical codes for covert communication. Even the ones that have proposed code designs have not provided any explicit optimal construction that achieves the limits predicted by information theory. We address this by developing a low-complexity coding scheme for covert communication using multilevel coding with pulse-position modulation. Our scheme achieves the covert capacity of discrete memoryless channels (DMCs) and additive white Gaussian noise (AWGN) channels. The rest of the thesis is organized as follows.

- In **Chapter 2**, we introduce the covert communication problem and the background required for this study. We also survey results in information-theoretic security that are relevant to covert communication.

- In **Chapter 3**, we introduce a coding scheme for covert communication over BI-DMCs using multilevel coding (MLC) and pulse-position modulation (PPM). We discuss the information-theoretic analysis of the scheme and prove that this scheme is optimal for covert communication.

- In **Chapter 4**, we discuss the design of practical codes for covert communication over BI-DMCs using MLC and PPM, and we provide an explicit low-complexity code construction using polar codes and invertible extractors.

- In **Chapter 5**, we extend the MLC-PPM code construction to covert communication over AWGN channels by proving the optimality of bi-orthogonal PPM as a signaling

scheme for AWGN channels.

- In **Chapter 6**, we demonstrate the application of MLC-PPM for covert forward reconciliation, which is an important step in covert secret-key generation.

- In **Chapter 7**, we provide some conclusions and perspectives.

CHAPTER 2
BACKGROUND: COVERT COMMUNICATION

In this chapter, we introduce the covert communication problem and review the related literature. The objective of covert communication is to establish communication between legitimate users while avoiding detection by an adversary. This is similar to the steganography discussed in [1], where a hidden message is embedded in an innocent message called *covertext* to get a seemingly innocent message called *stegotext*. In covert communication, we model the absence of communication by sending of "innocent symbols" by Alice and the presence of communication by sending of codewords which contain a mix of innocent symbols and non-innocent symbols. In steganography terms, the innocent symbols are the covertext and the codewords are the stegotext. The main difference between covert communication and steganography is that in covert communication, Bob and Willie observe Alice's codeword through a noisy channel whereas, in steganography, Bob and Willie have direct access to Alice's stegotext.

Covert communication also has some connections with spread-spectrum communication. Spread-spectrum communication techniques guarantee a low probability of interception and anti-jamming properties [2]. Interestingly, the development of spread-spectrum communication dates back to the Second World War, a war waged with jamming and anti-jamming tactics [2]. In spread-spectrum communication, instead of restricting a signal to a minimum bandwidth required by the signaling scheme, it is sent over a frequency band of a much larger order of magnitude either by frequency hopping techniques or spreading the signal using a pseudorandom spreading sequence. Although practical spread spectrum communication techniques are well established, there is no information-theoretic guarantee that it is not undetectable. Information-theoretically provable communication with low probability of detection (LPD) was first studied by Bash et. al. in [3]. We will devote

the rest of this chapter to introduce the notations and definitions and review the literature related to covert communication.

2.1 Notation

We denote random variables by uppercase letters, e.g., X, their realizations by lowercase, e.g., x, and vectors of random variables and their realizations by their corresponding boldface letters, e.g., $\mathbf{X}$ and $\mathbf{x}$, respectively. We denote a discrete memoryless channel with input $X \in \mathcal{X}$ and output $Y \in \mathcal{Y}$ with a transition probability distribution $W_{Y|X}$ by DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$. Given two distributions P and Q defined on the same alphabet $\mathcal{X}$, $\mathbb{V}(P, Q)$ denotes the total variation distance or the variational distance between the distributions and is defined as

$$\mathbb{V}(P, Q) \triangleq \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|, \tag{2.1}$$

and $\mathbb{D}(P\|Q)$ denotes the Kullback–Leibler(KL) divergence or the relative entropy between the same distributions and is defined as

$$\mathbb{D}(P\|Q) \triangleq \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{2.2}$$

Note that $\log$ denotes the natural logarithm and $\log_2$ denotes logarithm to the base 2.

If P and Q are defined on a continuous alphabet $\mathcal{X}$, then the variational distance is defined as

$$\mathbb{V}(P, Q) \triangleq \frac{1}{2} \int_{x \in \mathcal{X}} |P(x) - Q(x)| dx, \tag{2.3}$$

and the relative entropy is defined as

$$\mathbb{D}(P\|Q) \triangleq \int_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} dx. \tag{2.4}$$

Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be two discrete random variables with joint distribution P_{XY}. Let P_X be the marginal distribution of X and $P_{Y|X}$ be the conditional distribution of Y given X. Then, the entropy of X is defined as

$$H(X) \triangleq -\sum_{x \in \mathcal{X}} P_X(x) \log P_X(x). \tag{2.5}$$

The joint entropy of X and Y is defined as

$$H(XY) \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \log P_{XY}(x, y). \tag{2.6}$$

The conditional entropy of X given Y is defined as

$$H(Y|X) \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \log P_{Y|X}(y|x). \tag{2.7}$$

If X is a continuous random variable, the differential entropy is used instead of entropy and is defined as

$$h(X) \triangleq -\int_{x \in \mathcal{X}} P_X(x) \log P_X(x) dx. \tag{2.8}$$

The joint and conditional differential entropies are also defined similarly. The mutual information between X and Y is defined as

$$I(X; Y) \triangleq \mathbb{D}(P_{XY} \| P_X P_Y). \tag{2.9}$$

For the discrete case, we have $I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$, and for the continuous case, we have $I(X;Y) = h(X) - h(X|Y) = h(Y) - h(Y|X)$. Because we use natural logarithm to compute above quantities, we use nat as the unit of information.

For two real-valued functions $f(n)$ and $g(n)$ of $n \in \mathbb{N}$, we write $f(n) = o(g(n))$ if for all $\alpha > 0$, there exists $n_0 \in \mathbb{N}^*$ such that for all $n \geqslant n_0$, $|f(n)| \leqslant \alpha |g(n)|$; $f(n) = \mathcal{O}(g(n))$

if there exist $\alpha > 0$ and $n_0 \in \mathbb{N}^*$ such that for all $n \geqslant n_0$, $|f(n)| \leqslant \alpha|g(n)|$; $f(n) = \omega(g(n))$ if for all $\alpha > 0$, there exists $n_0 \in \mathbb{N}^*$ such that for all $n \geqslant n_0$, $|f(n)| \geqslant \alpha|g(n)|$.

2.2 Secret communication

A communication system is said to ensure secrecy (also known as confidentiality) if the legitimate users of the system can communicate with each other without revealing the content of their message to an adversary. In the layered communication systems that we use today, secrecy is achieved using encryption schemes (such as RSA and AES) in the higher layers of the protocol stack. However, these encryption schemes are only secure because the computational workload required to decrypt the message without the key is assumed to be beyond what is feasible with today's technology. In contrast, information-theoretically secure communication provides secrecy even if the adversary has infinite computational resources.

2.2.1 Shannon's cryptosystem

Shannon's cryptosystem model [4], illustrated in Fig. 2.1, initiated the information-theoretic analysis of secrecy. In this model, Alice encrypts a message W into a ciphertext $\mathbf{X}$ using a key S shared between Alice and Bob. Bob decodes the message from the ciphertext $\mathbf{X}$ using the key S. The decoded message is denoted by $\widehat{W}$. The eavesdropper Eve has ac-

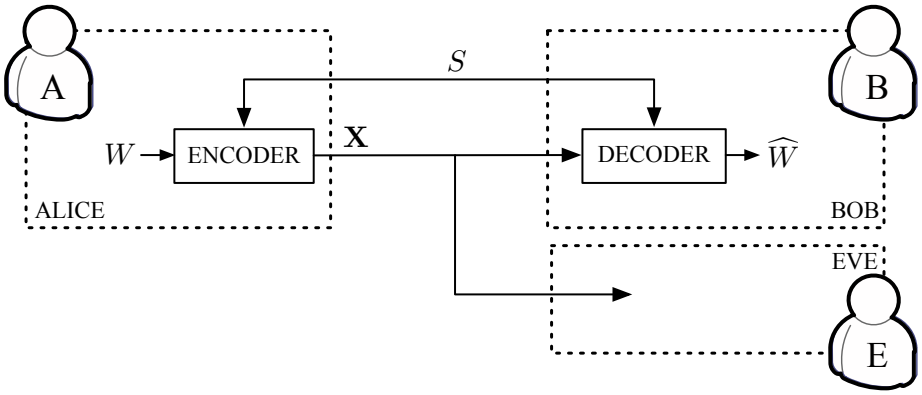

Figure 2.1: Shannon's cryptosystem.

cess to this ciphertext $\mathbf{X}$. Shannon introduced the notion of *equivocation* to measure the eavesdropper's uncertainty about the message as $H(W|\mathbf{X})$. A system is said to have *perfect secrecy* if $\widehat{W} = W$ and $H(W) = H(W|\mathbf{X})$, or equivalently $I(W;\mathbf{X}) = 0$. The quantity $I(W;\mathbf{X})$ can be interpreted as the amount of information *leaked* to the eavesdropper. The condition $I(W;\mathbf{X}) = 0$ implies that perfect secrecy is achieved if W and $\mathbf{X}$ are statistically independent. Unfortunately, perfect secrecy is only achievable with a one-time-pad and requires as many key bits as message bits.

2.2.2 Wiretap channel

In 1975, Wyner [5] introduced the wiretap channel, which is now the standard model to study information-theoretic secrecy over noisy channels. Unlike Shannon's model, the wiretap channel assumes that the communication channel is noisy. This assumption allows one to circumvent the rather disappointing result regarding Shannon's cryptosystem. The wiretap channel model, as illustrated in Fig. 2.2, consists again of legitimate users Alice and Bob, and an eavesdropper Eve. Alice first encodes her message W to a codeword $\mathbf{X}$ of length n using a coding scheme $\mathcal{C}_n$ and transmits the codeword over the DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ to Bob. Eve observes this communication over the DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$. Note that Alice and Bob do not have access to a shared secret key. Bob's and Eve's observations are denoted by $\mathbf{Y}$ and $\mathbf{Z}$, respectively, and the output of Bob's decoder is denoted by $\widehat{W}$. The coding scheme $\mathcal{C}_n$ is assumed to be known to Alice, Bob, and Eve. The reli-

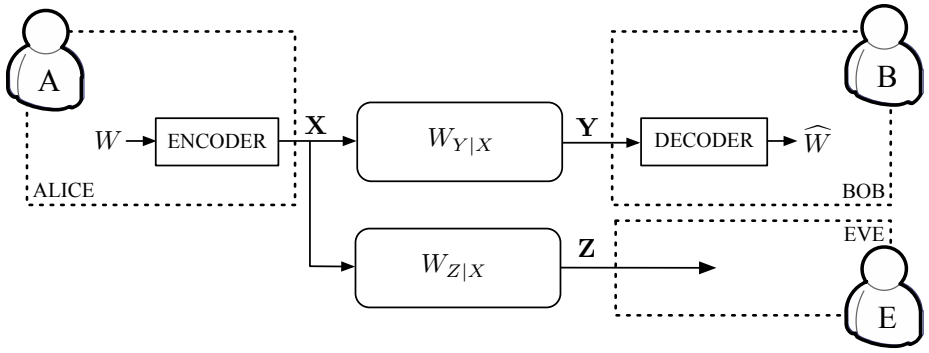

Figure 2.2: The wiretap channel model.

ability condition for the coding scheme $\mathcal{C}_n$ is expressed in terms of average probability of error and is given by

$$\lim_{n\to\infty} \mathbb{P}\Big(\widehat{W} \neq W|\mathcal{C}_n\Big) = 0.$$

Rather than enforcing perfect secrecy, the secrecy condition consists in requiring almost perfect secrecy. In the literature, there exists a variety of metrics to quantify what "almost" means. We discuss two types of secrecy metrics for which secrecy is measured in terms of the number of bits leaked to the adversary denoted by $I(W;\mathbf{Z})$. In *weak secrecy*, the $leakage$ is measured using the normalized mutual information, and the requirement is given by

$$\lim_{n\to\infty} \frac{1}{n} I(W;\mathbf{Z}|\mathcal{C}_n) = 0.$$

In *strong secrecy*, the leakage is measured using the unnormalized mutual information, and the requirement is given by

$$\lim_{n\to\infty} I(W;\mathbf{Z}|\mathcal{C}_n) = 0.$$

The maximum rate at which Alice can communicate with Bob while satisfying the reliability and secrecy requirements is called the *secrecy capacity*, denoted by C_S. The secrecy capacity is the same for weak and strong secrecy [6], and it is given by

$$C_S = \max_{V-X-YZ} [I(V;Y) - I(V;Z)]. \tag{2.10}$$

The notation $V - X - YZ$ means that V, X, Y, and Z form a Markov chain such that the joint distribution is $p_{VXYZ} = p_V p_{X|V} W_{YZ|X}$, and the maximization is taken over all joint distribution satisfying the Markov chain.

2.3 Covert communication

In its simplest form, the problem of covert communication (or communication with LPD), involves three users, Alice, Bob, and Willie, as shown in Figure 2.3. Alice wants to send

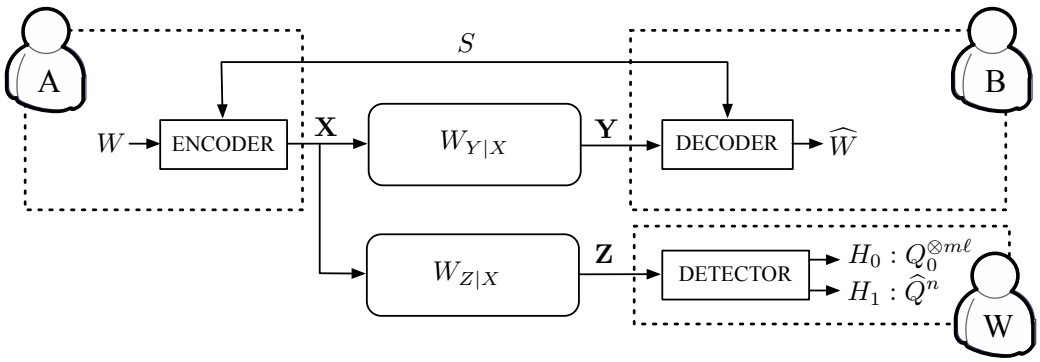

Figure 2.3: Covert communication model.

a message to Bob over a BI-DMC represented by DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ while avoiding the detection of the communication by the warden Willie, who observes the communication through DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$. By convention, channel input "0" is the innocent symbol corresponding to the absence of communication. For $j \in \{0, 1\}$ we set $P_j \triangleq W_{Y|X=j}$ and $Q_j \triangleq W_{Z|X=j}$. Alice encodes her messages $W \in [\![1, M]\!]$ into codewords $\mathbf{X}$ of n binary symbols, which are observed by Bob and Willie as $\mathbf{Y}$ and $\mathbf{Z}$, respectively. Alice's encoding may be assisted by private randomness $S \in [\![1, K]\!]$ shared only with Bob. Bob forms an estimate $\widehat{W}$ of the transmitted message using $\mathbf{Y}$ and S.

The objective of Alice is two-fold: 1) to communicate the message W to Bob with negligible probability of error and 2) to avoid detection by Willie. In this model, Willie knows the communication scheme used by Alice and Bob, and the only unknowns are the actual message transmitted and the realization of the channel noise. The probability of error at Bob's decoder is measured by $\mathbb{P}\left(W \neq \widehat{W}\right)$, where $\widehat{W}$ is Bob's estimate of the message W. From the perspective of Willie, the detection of communication involves choosing between two hypotheses – the hypothesis that Alice is not transmitting, denoted by H_0, and the hypothesis that Alice is transmitting, denoted by H_1. The performance of a hypothesis testing is characterized by the false alarm probability α, which is the probability that the detector outputs H_1 when H_0 is true, and missed detection probability β, which is the probability that the detector outputs H_0 when H_1 is true. Let $\widehat{Q}^n$ be the distributions induced when H_1 is true and $Q_0^{\otimes n}$ be the "innocent distribution" induced when H_0 is true.

Willie's objective is to minimize both α and β. However, α and β are correlated, and one cannot optimize them independently. The curve obtained by plotting α against $1-\beta$, known as ROC curve, captures this correlation. The worst performance of the detector corresponds to a blind test and Willie can achieve any $\alpha \in [0,1]$ and $\beta \in [0,1]$ such that $\alpha+\beta=1$ using a blind test. The ROC curve corresponding to this is depicted in Fig. 2.4.

From [7, Theorem 13.1.1], we have that the ROC curve for an optimal hypothesis test is characterized by

$$\alpha+\beta = 1-\mathbb{V}\Big(\widehat{Q}^n, Q_0^{\otimes n}\Big). \tag{2.11}$$

The variational distance can be bounded by relative entropy using Pinsker's inequality [8, Lemma 11.6.1] as

$$\mathbb{V}\Big(\widehat{Q}^n, Q_0^{\otimes n}\Big) \leqslant \sqrt{\frac{1}{2}\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)}. \tag{2.12}$$

By relaxing this inequality, we have $\mathbb{V}\Big(\widehat{Q}^n, Q_0^{\otimes n}\Big) \leqslant \sqrt{\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)}$, and hence, we obtain

$$\alpha+\beta \geqslant 1-\sqrt{\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)}, \tag{2.13}$$

Therefore, by using a coding scheme that ensures $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)$ to be very small, we can enforce the performance of Willie's detector to be no better than a blind test. In Fig. 2.4, the dashed line represents $\alpha+\beta = 1-\sqrt{\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)}$, and the right side of the dashed line represents $\alpha+\beta \geqslant 1-\sqrt{\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)}$. By ensuring $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)$ to be small, this dashed line moves closer to the line representing a blind test.

In this work, we use $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)$ as the measure for covertness. Note that the covertness condition, $\lim_{n\to\infty} \mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big) = 0$, does not ensure secrecy. It only ensures that the distribution induced in the presence of communication is close to the one in the absence

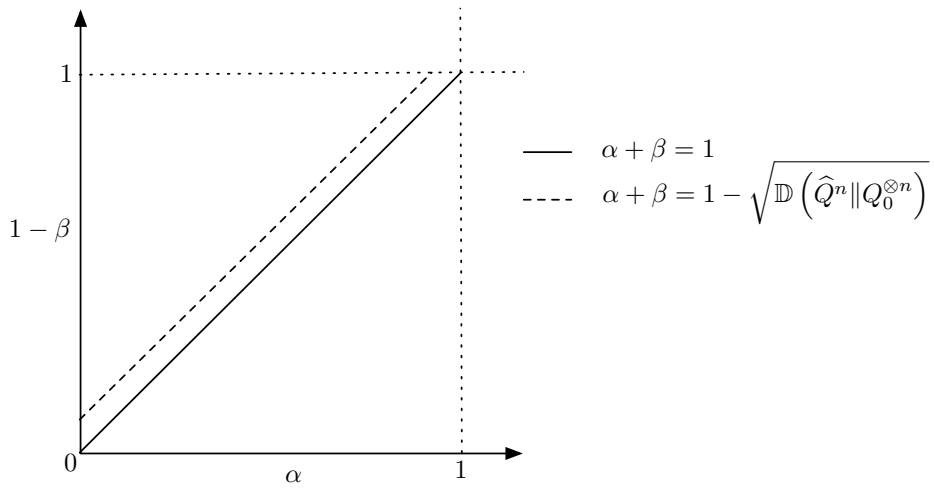

Figure 2.4: ROC curve for Willie's detector.

of communication. However, secrecy condition, $I(W;\mathbf{Z}|\mathcal{C}_n) \to 0$ as $n \to \infty$, requires the independence between the input W and the adversary's observation $\mathbf{Z}$, which is not implied from the covertness condition.

2.3.1 Previous works on covert communication

While practical covert communication techniques have been studied by the spread-spectrum community, the information-theoretic limits of covert communication were first studied by Bash et al. in [3]. In that work, the authors laid the foundations for the study of covert communication by discovering the square-root limit on the amount of information that can be transmitted reliably with LPD over AWGN channels. The authors analyzed a communication setup in which Alice transmits a real-valued symbol represented by the random variable X, and Bob and Willie observe the outputs of two AWGN channels represented by $Y = X + N_b$ and $Z = X + N_w$, respectively, where N_b and N_w are zero-mean Gaussian noise with variances σ_b^2 and σ_w^2. The achievability proof of covert communication over this channel uses a random coding argument in which codewords are generated independently and identically according to the distribution $p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} p_X(x_i)$, where $X \sim \mathcal{N}(0, P_f)$. The codebook is used for the transmission of a single message and is a secret shared between Alice and Bob; however, Willie knows how the codebook is generated. In the ab-

sence of transmission of codewords, the distribution at the output of Willie's channel is $Q_0^{\otimes n}$, where $Q_0 = \mathcal{N}(0, \sigma_w^2)$. Since Willie does not know the codebook, the distribution induced at the output of Willie's channel by the transmission of a codeword is $Q_1^{\otimes n}$, where $Q_1 = \mathcal{N}(0, P_f + \sigma_w^2)$. The authors showed that if Alice sets her average symbol power $P_f \leqslant 2\sqrt{2}\epsilon\frac{f(n)}{\sqrt{n}}$, the covertness condition can be satisfied by choosing $f(n) = \widehat{\sigma}_w^2$ if Alice knows σ_w is lower bounded by $\widehat{\sigma}_w$ and $f(n)$ such that $f(n) = o(1)$ and $f(n) = \omega(1/\sqrt{n})$ if σ_w is unknown to Alice. Further, the authors showed that Alice can reliably transmit, while satisfying the condition for covertness $\mathbb{V}(Q_0^{\otimes n}, Q_1^{\otimes n}) \leqslant \epsilon$ for any $\epsilon > 0$, $o(\sqrt{n})$ bits to Bob over n channel uses when σ_w is unknown to Alice and $O(\sqrt{n})$ bits over n channel uses if a lower bound $\sigma_w \geqslant \widehat{\sigma}_w$ for some $\widehat{\sigma}_w > 0$ is known to her. Moreover, the authors showed that if Alice attempts to transmit $\omega(\sqrt{n})$ bits over n channel uses, either Willie can detect the communication using a power detector or Bob cannot decode the codeword with an arbitrarily low probability of error.

Subsequently, Che et al. [9] extended the result to BSCs by establishing the square root limit without the need for any shared secret keys when the channels are BSCs and Willie's observations are significantly noisier than Bob's. Bloch [10] generalized the results to all DMCs and AWGN channels showing that communication with LPD is possible without secret key as long as the legitimate receiver's channel is "better" than the warden's channel in the sense that $\mathbb{D}(P_1\|P_0) > \mathbb{D}(Q_1\|Q_0)$. In [11], Bash et al. showed that selecting a single n-symbol-period slot t_A out of $T(n)$ slots, as depicted in Fig. 2.5, using $\log T(n)$ pre-shared secret bits allows Alice to reliably transmit $\mathcal{O}(\sqrt{n \log T(n)})$ bits on an AWGN channel to Bob with LPD for Willie's detector.

Covert communication also relates to the problem of stealth, introduced by Hou and

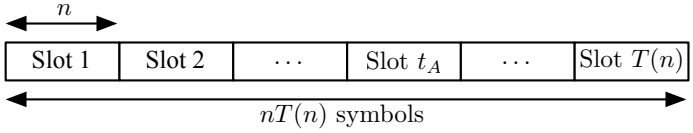

Figure 2.5: Slotted covert transmission.

Kramer [12]. Hou and Kramer combined secrecy and covertness by defining a security measure called "effective security" that includes strong secrecy and stealth communication as

$$\mathbb{D}(P_{M\mathbf{Z}}\|P_M Q_{\mathbf{Z}}) = I(M;\mathbf{Z}) + \mathbb{D}(P_{\mathbf{Z}}\|Q_{\mathbf{Z}}), \tag{2.14}$$

where M is the message with distribution P_M, $\mathbf{Z}$ is the adversary's observation with distribution $P_{\mathbf{Z}}$ and joint distribution $P_{M\mathbf{Z}}$, and $Q_{\mathbf{Z}}$ is the distribution that the warden expects to observe when the transmitter is not communicating useful messages. Ensuring $\mathbb{D}(P_{M\mathbf{Z}}\|P_M Q_{\mathbf{Z}}) \to 0$ as $n \to \infty$ implies the coding scheme satisfies the secrecy condition $I(M;\mathbf{Z}) \to 0$ and the stealth condition $\mathbb{D}(P_{\mathbf{Z}}\|Q_{\mathbf{Z}}) \to 0$. A subtle distinction of covert communication is that $Q_{\mathbf{Z}} = Q_0^{\otimes n}$, a distribution induced by a deterministic input. In [13], Che et al. discussed a reliable, deniable, and hidable communication scheme over a BSC. Che et al. showed that the effective secrecy, which constitutes deniability (covertness) and hidability (secrecy), could be achieved by encoding for each separately. In [14], Kadhe et al. analyzed a reliable, deniable, and hidable communication over a multipath network consisting of multiple parallel links in the presence of a passive warden Willie, who observes an unknown subset of links. The authors also characterized the capacity for reliable and deniable communication for that network. A survey of various security measures related to covert communication can be found in [15].

The square root limit on the number of bits transmitted implies that covert communication is only possible at *zero* rate. To obtain a meaningful asymptotic constant, Wang et al. [16] characterized the asymptotic constant obtained by normalizing the number of bits transmitted with $\sqrt{n\delta}$ such that $\mathbb{D}\left(\widehat{Q}^n\|Q_0^{\otimes n}\right) \leqslant \delta$. This quantity is now known as the covert capacity in the literature and will be precisely defined in Section 2.3.2. Bloch [17] provided a complete characterization of the covert capacity with the amount of secret key required and determined the condition under which keyless covert communication is pos-

sible. Later, Tahmasbi and Bloch [18] analyzed covertness using three different metrics – the relative entropy and the variational distance between the induced output distribution and innocent distribution and the probability of missed detection for a fixed probability of false alarm. For each metric, the authors characterized the exact first-order asymptotics of the number of bits that can be transmitted for a given maximum probability of error and covertness metric. The authors also characterized the exact second-order asymptotics for relative entropy and derived bounds for the other two metrics. Meanwhile, Arumugam et al. extended the information-theoretic analysis of covert communication to multiple-access channels (MACs) [19, 20], broadcast channels [21, 22], and relay channels [23]. Moreover, the idea of covertness has been extended to covert secret key generation in [24] and the quantum regime in [25, 26, 27, 28, 29].

Several studies have investigated the possibility of circumventing the square-root law by incorporating the warden's uncertainty about the channel noise parameters. In one study [30], the authors analyzed a covert communication setting in which the channels are BSCs with the probability of bit-flipping distributed uniformly in a known interval and Willie's channel is noisier than Bob's channel and showed that positive-rate covert communication is possible in that case. In another study, Lee et al. [31] analyzed undetectable communication over AWGN and Rayleigh channels when the warden employs a radiometer detector and has uncertainty about his noise variance. The authors showed that positive-rate covert communication is possible in that scenario. The same problem was extended in [32] to multiple-input multiple-output (MIMO) Rayleigh fading channels. Further in [33], the authors analyzed the effect of sampling of Willie's observations. In another study, Sobers et. al. [34] analyzed a covert communication scenario in the presence of a friendly jammer that does not know the content or timing of Alice's transmission and showed that a power detector is optimal for Willie when Alice uses a random Gaussian codebook. Applying that result, the authors showed that Alice can communicate $\mathcal{O}(n)$ covert bits to Bob in n channel uses. In our research, we will be mainly concerned with situations in which the

square root law cannot be circumvented.

2.3.2 Covert capacity

In our covert communication model in Fig. 2.3, the message W is selected from the set $[\![1, M]\!]$ and the key S is selected from the set $[\![1, K]\!]$. Hence, $\log M$ represents the number of $nats$ of message transmitted and $\log K$ represents the number of $nats$ of key used for the transmission. As mentioned in the previous section, the scaling of the number of bits transmitted in n channel uses follows square root law. Hence, the rate of communication in the usual sense of the number of bits transmitted per channel use is zero. From the characterization of the constant behind the $\sqrt{n}$ in [16] for the maximum number of bits that can be transmitted covertly over a DMC such that $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big) \leqslant \delta$ for $\delta > 0$ and the probability of error goes to zero, Wang et. al. noticed that a constant that depends only on the channel properties can be obtained by normalizing the number of bits transmitted by $\sqrt{n\delta}$. Using that characterization, we define covert capacity as follows.

As explained previously, we shall measure the reliability of Bob's decoder using the probability of error $\mathbb{P}\Big(W \neq \widehat{W}\Big)$ and the covertness for Willie's detector using the relative entropy $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big)$. We say that a coding scheme achieves covert throughput R with a covert key throughput R_K if there exists a sequence of codes with block length n such that

$$\lim_{n\to\infty} \frac{\log M}{\sqrt{n\delta}} \geqslant R, \qquad \lim_{n\to\infty} \frac{\log K}{\sqrt{n\delta}} \leqslant R_K,$$
$$\lim_{n\to\infty} \mathbb{P}\Big(W \neq \widehat{W}\Big) = 0, \text{ and } \lim_{n\to\infty} \mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big) \leqslant \delta,$$

for some chosen $\delta > 0$. The supremum of all achievable covert throughputs is called the covert capacity and has been characterized in [16, 17] and [18, Theorem 1] as

$$C_{\text{covert}} = \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}}\mathbb{D}(P_1\|P_0). \tag{2.15}$$

In [18], the authors refined the previous result and analyzed the first- and second-order asymptotics of the maximum number of bits that can be transmitted over a DMC by considering three covertness metrics. The reliability and covertness conditions the authors considered for covert communication are that the maximum probability of error should be upper bounded by ϵ and the covertness metric should be upper bounded by δ. When the covertness condition is $\mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big) \leqslant \delta$, the optimal number of bits [18, Theorem 1] is given by

$$\log M = \sqrt{\frac{2\delta}{\chi_2(Q_1||Q_0)}} D_P \, n^{\frac{1}{2}} - \sqrt{\sqrt{\frac{2\delta}{\chi_2(Q_1||Q_0)}} V_P} \, Q^{-1}(\epsilon) \, n^{\frac{1}{4}} + \mathcal{O}(\log n), \quad (2.16)$$

where $D_P \triangleq \mathbb{D}(P_1 \| P_0)$, $V_P \triangleq \text{Var}\Big(\log \frac{P_1(Y)}{P_0(Y)} | P_1\Big)$, and Q^{-1} denotes the inverse of the Q-function $Q(x) \triangleq \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$. Note that we obtain the covert capacity in (2.15) from the above equation by taking $\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{\log M}{\sqrt{n\delta}}$.

When the covertness condition is $\mathbb{V}\Big(\widehat{Q}^n, Q_0^{\otimes n}\Big) \leqslant \delta$, the authors obtained the exact first order asymptotic for the optimal number of bits [18, Theorem 2] given by

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{\log M}{\sqrt{n}} = \frac{2\Gamma D_P}{\sqrt{\chi_2(Q_1||Q_0)}}, \quad (2.17)$$

where $\Gamma \triangleq Q^{-1}\left(\frac{1-\delta}{2}\right)$. The authors also obtained bounds for the asymptotic of the second order term.

When the covertness condition is expressed in terms of optimal probability of missed detection $\beta_\alpha\left(\widehat{Q}^n, Q_0^{\otimes n}\right)$ for a given probability of false alarm α, the authors obtained the exact first order asymptotic for the optimal number of bits while satisfying the covertness condition $\beta_\alpha\left(\widehat{Q}^n, Q_0^{\otimes n}\right) \geqslant 1 - \alpha - \delta$. The first order asymptotic [18, Theorem 3] is given by

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \frac{\log M}{\sqrt{n}} = \frac{(\Lambda + \Upsilon) D_P}{\sqrt{\chi_2(Q_1||Q_0)}}, \quad (2.18)$$

where $\Lambda \triangleq Q^{-1}(1-\alpha-\delta)$ and $\Upsilon \triangleq Q^{-1}(\alpha)$. The authors also obtained bounds for the second-order asymptotic.

In this work, we consider relative entropy as the covertness metric. For this case, the minimum covert key throughput needed to achieve the covert capacity is denoted by R_K^{opt}. From [18, Theorem 1], we have

$$R_K^{\text{opt}} = \sqrt{\frac{2}{\chi_2(Q_1\|Q_0)}}\left(\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0)\right)^+, \tag{2.19}$$

where $(x)^+ \triangleq \max(x, 0)$. In this work, we design codes that achieve this optimal throughputs over DMCs and AWGN channels using multilevel coding (MLC) and pulse-position modulation (PPM) with multistage decoding (MSD).

2.4 Achieving covertness with resolvability codes

Resolvability codes are used to approximate the output distribution of a channel by employing a codebook at the input of the channel. Han and Verdú [35] defined the notion of resolvability of a channel as "the number of random bits per input sample required to achieve arbitrarily accurate approximation of the output statistics regardless of the actual input process." For a DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$, let P_X denote the input distribution and P_Y denote the corresponding output distribution. i.e.,

$$P_Y(y) = \sum_{x\in\mathcal{X}} P_X(x) W_{Y|X}(y|x). \tag{2.20}$$

Let $\mathbf{X}$ denote an input sequence of length n and $\mathbf{Y}$ denote the corresponding output. Let $P_{\mathbf{X}}^{\otimes n}$ and $P_{\mathbf{Y}}^{\otimes n}$ denote the product distributions for n uses of the channel. Let $\mathcal{C}_n = \{\mathbf{x}_i\}_{i\in[\![1,2^{nR}]\!]}$ denote a code with each codeword of length n. Let $\widehat{P}_{\mathbf{Y}}$ denote the output

distribution induced by sending a codeword selected uniformly at random from $\mathcal{C}_n$. i.e.,

$$\widehat{P}_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2^{nR}} \sum_{i=1}^{2^{nR}} W_{Y|X}^{\otimes n}(\mathbf{y}|\mathbf{x}_i). \tag{2.21}$$

We say that R is an achievable resolution rate if there exists a sequence of codes $\{\mathcal{C}_n\}$ such that $\lim_{n\to\infty} \mathbb{D}\Big(\widehat{P}_{\mathbf{Y}} \| P_{\mathbf{Y}}^{\otimes n}\Big) = 0$. The minimum of all achievable resolution rate is called the resolvability of the channel. Han and Verdú [35] showed that the resolvability is equal to the Shannon capacity $I(X;Y)$.

The importance of resolvability based code construction for achieving secrecy was shown in [36], where the objective was to control the distribution at the eavesdroppers channel output such that the observations would be independent of the message transmitted. In covert communication, the objective is to induce a distribution $P_{\mathbf{Z}}$ at the warden's channel output such that it is indistinguishable from the innocent distribution $Q_0^{\otimes n}$. Bloch [17] proposed achievability schemes for covert communication using resolvability in two steps. First, distribution $Q_{\alpha_n}^{\otimes n}$ induced when the input of the channel is biased such that the probability of symbol 1 is $\alpha_n \triangleq \frac{\omega_n}{\sqrt{n}}$ with $\omega_n = o(1) \cap \omega(1/\sqrt{n})$ is shown to be indistinguishable from $Q_0^{\otimes n}$, i.e., $\lim_{n\to\infty} \mathbb{D}\big(Q_{\alpha_n}^{\otimes n} \| Q_0^{\otimes n}\big) = 0$. Then, resolvability based coding schemes are used to simulate the distribution $Q_{\alpha_n}^{\otimes n}$.

Note, however, that the approach in [17] is still information-theoretic in nature and does *not* offer explicit low-complexity algorithms. Few studies have explored the explicit construction of codes for resolvability. In one study [37], the authors presented a polar coding based code for resolvability of symmetric BI-DMC as the basis for a construction of code for strong coordination. Subsequently, Chou et al. generalized the method for any BI-DMCs for construction of codes for empirical and strong coordination [38] and broadcast channel with confidential messages [39]. In another work [40], Chou et al. presented a construction of low-complexity channel resolvability codes for symmetric MAC using invertible extractors and injective group homomorphisms. Bloch et al. presented a

comprehensive overview of coding schemes for secrecy with approaches based on channel resolvability, channel reliability, and source coding in [6].

2.5 Codes for covert communication

Although there is a significant amount of literature on the information-theoretic analysis of covert communication, few researchers have addressed the construction of practical low-complexity codes. In [41], the authors analyzed a concatenated code construction with an inner random code and an outer Reed-Solomon code and showed the existence of low-complexity coding schemes for covert communication over a BSC, but their work did not provide an explicit code construction of the inner code. In another study [42] that we will review later, the authors proposed an explicit code construction for asynchronous covert communication [43] over BI-DMCs using polar codes, which involves hiding the transmission in a larger window using a shared secret key. In [44], the authors showed the optimality of PPM for covert communication by presenting an achievability proof using a random non-binary code over PPM symbols. In our work, we build upon this construction to achieve a low-complexity coding scheme that achieves the covert capacity for BI-DMCs. We will review the main results on PPM based construction in the next section.

2.5.1 Polar codes for covert communications over asynchronous discrete memoryless channels

In [42], the authors introduced an explicit code for covert communication over binary-input asynchronous discrete memoryless channels using binary polar codes. This code construction is based on the coding scheme for asynchronous covert communication using random codes proposed in [43]. The coding scheme for asynchronous covert communication is as follows. The sender encodes an uniformly distributed message $W \in [\![1, M_n]\!]$ into a codeword of length n, possibly with the help of secret key $S \in [\![1, K_n]\!]$ only known to Alice and Bob. The codebook is public and known to all parties. The scheme is called "asyn-

chronous" because the codeword is embedded within a larger transmission window of size $N > n$ by choosing the starting index T of the codeword uniformly at random between 1 and $N' \triangleq N - n + 1$, where N is a function of n. We refer to the part of the transmission window where the codeword is transmitted as the codeword window. Let $X_{1:N}$ represent the sequence transmitted during the transmission window and $Y_{1:N}$ and $Z_{1:N}$ represent the sequences observed by Bob and Willie, respectively. Let $\widehat{Q}^n$ denote the distribution in-

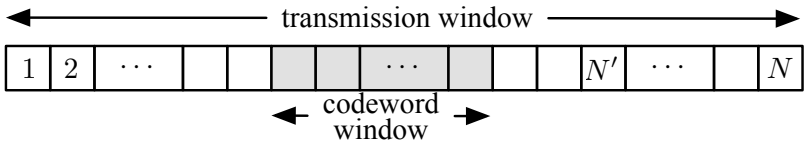

Figure 2.6: Illustration of transmission window for asynchronous covert communication.

duced at the output of Willie's channel by the codeword window. The distribution induced at the output of Willie's channel by a transmission window with codeword window starting at a known index t can be expressed as

$$\widehat{Q}_t^N(\mathbf{z}) = \prod_{k=1}^{t-1} Q_0(z_k) \prod_{k=t}^{t+n-1} \widehat{Q}^n(z_k) \prod_{k=t+n}^{N} Q_0(z_k). \tag{2.22}$$

Then, the distribution induced at Willie's channel by randomizing the start index of codeword window in a transmission window is denoted by $\widehat{Q}^N$ and is given by

$$\widehat{Q}^N \triangleq \mathbb{E}_T\left[\widehat{Q}_T^N\right]. \tag{2.23}$$

Bob estimates the message W from the observation $Y_{1:n}$ and the secret key S. The estimated message is denoted by $\widehat{W}$ and the performance of Bob's decoder is measured by $P_e^{(n)} \triangleq \mathbb{E}_S\left[\mathbb{P}\left(\widehat{W} \neq W|S\right)\right]$. The covertness of the transmission is measured by the total variation between the induced distribution $\widehat{Q}^N$ and the innocent distribution $Q_0^{\otimes N}$ denoted by $V^{(n)} \triangleq \mathbb{V}\left(\widehat{Q}^N, Q_0^{\otimes N}\right)$. The main result of [43] regarding the existence of codes for asynchronous covert communication is given by the following proposition.

Proposition 1. *[42, Proposition 1] Consider sequences of positive numbers* $\{\alpha_n\}_{n\in\mathbb{N}^*}$, $\{\beta_n\}_{n\in\mathbb{N}^*}$ *such that* $\alpha_n \in \omega(\frac{1}{\sqrt{n}}) \cap o(1)$, $\beta_n = \omega\left(2^{-\frac{n\alpha_n}{\log n}}\right) \cap o(1)$ *as* n *goes to infinity. For* $N = \frac{2^{n\alpha_n^2}}{\beta_n \alpha_n^2}$, *there exist sequences of codes of block length* n *hidden in transmission windows of size* N *such that*

$$\lim_{n\to\infty} \frac{\log M_n}{n\alpha_n} \geqslant \mathbb{D}(P_1\|P_0) \tag{2.24}$$

$$\lim_{n\to\infty} \frac{\log K_n}{n\alpha_n} \leqslant [\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0)]^+ \tag{2.25}$$

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{2.26}$$

$$\lim_{n\to\infty} V^{(n)} = 0. \tag{2.27}$$

The proof of the above proposition uses a random coding argument [43]. The work in [42] establishes a similar result using polar codes in place of random codes by a slight modification of the above coding scheme by considering b_n consecutive transmission windows of size N. The probability of error $P_e^{(n)}$ for Bob's decoder is modified to reflect the set of messages in b_n transmission windows $\{W_i\}_{i=1}^{b_n}$ as

$$P_e^{(n)} = \mathbb{P}\Big(\{\widehat{W}_i\}_{i=1}^{b_n} \neq \{W_i\}_{i=1}^{b_n}\Big), \tag{2.28}$$

and the covertness is measured by the total variation between the distribution induced over b_n transmission windows given by

$$V^{(n)} = \mathbb{V}\Big(\widehat{Q}^{b_n N}, Q_0^{\otimes b_n N}\Big). \tag{2.29}$$

The results of [42] is summarized in the following proposition.

Proposition 2. *[42, Proposition 2] There exists a constant* $\kappa \in]0, \frac{1}{2}[$ *such that for all sequences of positive numbers* $\{\alpha_n\}_{n\in\mathbb{D}} \in \omega(\frac{1}{n^\kappa}) \cap o(\frac{1}{\log n})$, $\{\beta_n\}_{n\in\mathbb{D}} \in \omega(2^{-\frac{n\alpha_n}{\log n}}) \cap$

$o(\frac{1}{\log n})$*and sequence of integers* $\{b_n\}_{n\in\mathbb{D}} \in \omega(\log n) \cap o(\frac{1}{\beta_n}) \cap o(n)$ *as* n *goes to infinity, there exist low-complexity polar-code based schemes operating over* b_n *transmission windows of size* $N = \frac{2^{n\alpha_n^2}}{\beta_n \alpha_n^2}$*, each embedding a codeword window of length* n*, with*

$$\lim_{n\to\infty} \frac{\log M_n}{n b_n \alpha_n} \geqslant \mathbb{D}(P_1\|P_0) \tag{2.30}$$

$$\lim_{n\to\infty} \frac{\log K_n}{n b_n \alpha_n} \leqslant [\mathbb{D}(Q_1\|Q_0) - \mathbb{D}(P_1\|P_0)]^+ \tag{2.31}$$

$$\lim_{n\to\infty} P_e^{(n)} = 0 \tag{2.32}$$

$$\lim_{n\to\infty} V^{(n)} = 0. \tag{2.33}$$

The constant κ arises from the analysis of fine polarization of vanishing entropy rate sources [42, Proposition 3] and the authors noted that κ takes values of the order of 10^{-3}. This means that the range of values α_n takes is limited and it cannot achieve the scaling of $\frac{1}{\sqrt{n}}$ required to achieve the scaling of Proposition 1.

2.5.2 PPM for covert communication

A PPM symbol of order m is a vector of length m with symbol 0 at $m-1$ positions and symbol 1 at one position. Let $\widetilde{x}_i$ denote the vector with symbol 1 at i-th component as shown in Fig. 2.7, and let $\widetilde{\mathcal{X}}_m = \{\widetilde{x}_i\}_{i\in[\![1,m]\!]}$ denote the set of all possible PPM symbols of order m. Let $\widetilde{Z} = (Z_1, Z_2, \ldots, Z_m)$ denote the output of the channel DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ corresponding to a vector input of length m. Also, let Q_{PPM}^m be the output distribution induced by an input uniformly distributed on $\widetilde{\mathcal{X}}_m$ over a DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ and $Q_0^{\otimes m}$ be the output distribution induced by an all-zero input. For a realization of output of the

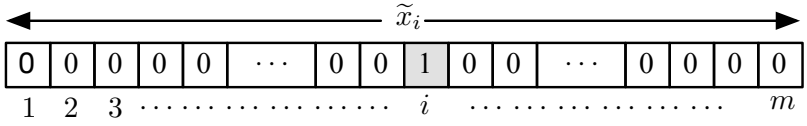

Figure 2.7: Illustration of PPM symbol $\widetilde{x}_i$.

channel $\widetilde{z} = (z_1, z_2, \ldots, z_m)$, we have

$$Q_0^{\otimes m}(\widetilde{z}) = \prod_{i=1}^{m} W_{Z|X}(z_i|0),$$

and

$$Q_{\text{PPM}}^{m}(\widetilde{z}) = \frac{1}{m}\sum_{i=1}^{m} W_{Z|X}^{\otimes m}(\widetilde{z}|\widetilde{x}_i) = \frac{1}{m}\sum_{i=1}^{m} Q_0^{\otimes m}(\widetilde{z})\frac{Q_1(z_i)}{Q_0(z_i)}.$$

The optimality of PPM for covert communication over BI-DMCs follows from the sparse nature of PPM symbols. Because of the sparse structure of PPM symbols, we can bound on the relative entropy between the two distributions Q_{PPM}^{m} and $Q_0^{\otimes m}$ as follows.

Lemma 1. *[44, Lemma 1] The relative entropy between the distributions Q_{PPM}^{m} and $Q_0^{\otimes m}$ defined above satisfies*

$$\mathbb{D}(Q_{\text{PPM}}^{m}\|Q_0^{\otimes m}) \leqslant \frac{1}{2m}\chi_2(Q_1||Q_0) + \frac{1}{m^2}\left(\chi_2(Q_1||Q_0)^2 - \frac{1}{6}\chi_3(Q_1||Q_0)\right) + \frac{1}{m^3}\left(\frac{1}{3}\chi_4(Q_1||Q_0) - \chi_2(Q_1||Q_0)^2\right), \quad (2.34)$$

where $\chi_k(Q_1||Q_0) \triangleq \sum_{z\in\mathcal{Z}} \frac{(Q_1(z)-Q_0(z))^k}{Q_0(z)}$ for any integer $k \geqslant 1$.

Note that the original version of above lemma in [44, Lemma 1] has different multipliers for the $\frac{1}{m^2}$ and $\frac{1}{m^3}$ terms stemming from an error in the calculation. However, it does not affect the other results in [44], and we have corrected that error in the present lemma.

Lemma 1 shows that the two distributions can be made indistinguishable by scaling the order m of the PPM symbols. Let $\widetilde{X}$ denote the random variable uniformly distributed in $\widetilde{\mathcal{X}}_m$. Let $\widetilde{Y}$ denote the random variable representing the output of the channel $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ corresponding to the input $\widetilde{X}$.

Lemma 2. *[44, Lemma 2] The mutual information between $\widetilde{X}$ and $\widetilde{Y}$ satisfies*

$$\mathbb{D}(P_1\|P_0) - \frac{1}{m}\chi_2(P_1||P_0) \leqslant I\left(\widetilde{X};\widetilde{Y}\right) \leqslant \mathbb{D}(P_1\|P_0). \quad (2.35)$$

The above lemma suggests that it is possible to transmit close to $\mathbb{D}(P_1\|P_0)$ bits in m uses of the channel $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ by coding over many PPM symbols. In [44], the authors showed the optimality of PPM for covert communication using a random coding scheme over PPM symbols by appropriately scaling the order of the PPM symbols proportionally to the codeword length, which is summarized by the following lemma.

Proposition 3. *[44, Proposition 1] For any $\xi, \delta > 0$, there exists a sequence of m-ary codes of length $\ell \triangleq \left\lceil \frac{2\delta}{\chi_2(Q_1||Q_0)} m \right\rceil$ over PPM of order m, with effective blocklength $n \triangleq \ell m$, such that*

$$\begin{aligned}
\lim_{n\to\infty} \frac{\log M}{\sqrt{n\delta}} &\geqslant (1-\xi)\sqrt{\frac{2}{\chi_2(Q_1||Q_0)}}\mathbb{D}(P_1\|P_0),\\
\lim_{n\to\infty} \frac{\log MK}{\sqrt{n\delta}} &\leqslant (1+\xi)\sqrt{\frac{2}{\chi_2(Q_1||Q_0)}}\mathbb{D}(Q_1\|Q_0),\\
\lim_{n\to\infty} P_e^{(n)} = 0, &\quad \lim_{n\to\infty} \mathbb{D}\Big(\widehat{Q}^n \| Q_0^{\otimes n}\Big) \leqslant \delta,
\end{aligned}$$

where $\widehat{Q}^n$ is the output distribution induced by the coding scheme.

Note that the results in [44] show the existence of codes using random coding argument. In our work, we propose an explicit low-complexity scheme using MLC over PPM symbols.

2.6 Binary linear codes cannot achieve the covert capacity of BI-DMCs without shared secret

We now consider the model described in Section 2.3 with Bob's channel represented by BI-DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ and Willie's channel represented by BI-DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$, where $\mathcal{X} \triangleq \{0, 1\}$ and 0 is the innocent symbol for covert communication corresponding to the absence of transmission. According to the results in [17], covert communication is possible without any secret key shared between Alice and Bob if $\mathbb{D}(P_1\|P_0) > \mathbb{D}(Q_1\|Q_0)$. Our

objective is to show that, under this assumption, covert communication with binary linear codes is not possible without shared secret key because such communication is easily detected; in particular, we show that no binary linear code can achieve the covert capacity.

A small misconception One might think that linear codes might not be covert because the linearity constraint requires the sum of codewords to be a codeword; consequently, linear combinations of low-weight codewords would result in a codeword with high weight. This is, however, insufficient to argue that linear codes cannot be covert. In fact, one could consider a systematic code with *unit-weight* codewords in the generator matrix, i.e, $\mathbf{G} = (\ \mathbf{I}_k \quad \mathbf{0}_{k\times n-k}\ )$, where $\mathbf{I}_k$ is the identity matrix of size k. All codewords of this code have weight at most k. Of course, this code would hardly be covert, mainly because the *structure* of the generator matrix allows an attacker to dismiss the last $n-k$ codeword components. The next proposition formalizes this.

Proposition 4. *Consider an (n,k) binary linear code with $(n,k) \in (\mathbb{N}^*)^2, n \geqslant k$. Assume that the code is used for communication over a BI-DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ with uniformly distributed messages. There exists a binary hypothesis test with false alarm probability α and missed-detection probability β such that*

$$
\begin{aligned}
0 \leqslant \alpha &\leqslant \frac{16}{k\chi_2(Q_1\|Q_0)},\\
0 \leqslant \beta &\leqslant \frac{1}{k}\left(\frac{16}{\chi_2(Q_1\|Q_0)} + \frac{8\chi_3(Q_1\|Q_0)}{\chi_2(Q_1\|Q_0)^2} - 4\right).
\end{aligned}
$$

Proof. Let $\mathbf{G} \in \mathbb{F}_2^{k\times n}$ be a generator matrix of the code, with columns $\{\mathbf{g}_i\}_{i=1}^n \in \mathbb{F}_2^k$. Denote the set of 2^k codewords by $\mathcal{C} \triangleq \{\mathbf{c}_\ell\}_{\ell=1}^{2^k}$. Let m be the number of non all-zero columns in $\mathbf{G}$, and denote the corresponding column indices $\{i_j\}_{j=1}^m$. Note that $k \leqslant m$ by definition of the dimension of the code. The observations of the adversary in the m positions $\{i_j\}_{j=1}^m$ constitute a sufficient statistics for the detection of communication, and the optimal test is a log-likelihood ratio test restricted to the m positions. We consider a

suboptimal test that only operates on a subset of indices $\mathcal{S} \subset \{i_j\}_{j=1}^m$ that indexes all the distinct columns of $\mathbf{G}$. Since the matrix $\mathbf{G}$ is binary, this implies that the columns $\{\mathbf{g}_i\}_{i\in\mathcal{S}}$ are pairwise linearly independent and $|\mathcal{S}| \geqslant k$. Given an observation $\mathbf{z} = (z_1, \ldots, z_n)$ at the output of the channel, construct the statistics

$$T(\mathbf{z}) \triangleq \frac{1}{|\mathcal{S}|} \sum_{j\in\mathcal{S}} \frac{Q_1(z_j) - Q_0(z_j)}{Q_0(z_j)}. \tag{2.36}$$

We now show that there exists a test on this statistic that satisfy the conditions in the proposition. We use Markov's inequality to show this. Towards this end, we first compute the mean and variance of $T(\mathbf{Z})$ in the absence and the presence of communication.

In the absence of communication, we have for any $j \in \mathcal{S}$ that $Z_j \sim Q_0$, and therefore

$$\mathbb{E}_{Q_0^{\otimes n}}[T(\mathbf{Z})] = \frac{1}{|\mathcal{S}|} \sum_{j\in\mathcal{S}} \mathbb{E}_{Q_0}\left[\frac{Q_1(Z_j) - Q_0(Z_j)}{Q_0(Z_j)}\right] = 0, \tag{2.37}$$

$$\begin{aligned}\mathrm{Var}_{Q_0^{\otimes n}}\left(T(\mathbf{Z})\right) = \frac{1}{|\mathcal{S}|^2}\Bigg(&\sum_{j\in\mathcal{S}} \mathbb{E}_{Q_0}\left[\frac{(Q_1(Z_j) - Q_0(Z_j))^2}{Q_0(Z_j)^2}\right] \\ &+2 \sum_{j,\ell\in\mathcal{S}:j<\ell} \mathbb{E}_{Q_0}\left[\frac{(Q_1(Z_j) - Q_0(Z_j))}{Q_0(Z_j)}\right]\mathbb{E}_{Q_0}\left[\frac{(Q_1(Z_\ell) - Q_0(Z_\ell))}{Q_0(Z_\ell)}\right]\Bigg)\end{aligned} \tag{2.38}$$

$$= \frac{1}{|\mathcal{S}|}\chi_2(Q_1\|Q_0). \tag{2.39}$$

In the presence of communication, let $\widehat{Q}^n$ represent the output distribution induced by the code, where

$$\widehat{Q}^n(\mathbf{z}) = \frac{1}{2^k} \sum_{\mathbf{v}\in\mathbb{F}_2^k} W_{Z|X}^{\otimes n}(\mathbf{z}|\mathbf{v}\mathbf{G}). \tag{2.40}$$

For any $j \in \mathcal{S}$, we have $Z_j \sim \frac{1}{2}Q_0 + \frac{1}{2}Q_1$ by [45, Problem 3.25]. Consider now $j, \ell \in \mathcal{S}$, $j \neq \ell$ and define $N_{ab} = |\{\mathbf{c} = (c_1, \ldots, c_n) \in \mathcal{C} : c_j = a, c_\ell = b\}|$ for $a, b \in \{0, 1\}$. By definition of $\mathcal{S}$, columns $\mathbf{g}_j$ and $\mathbf{g}_\ell$ are linearly independent so that $N_{00} = N_{01} = N_{10} =$

$N_{11} = 2^{k-2}$. Hence,

$$\mathbb{P}(Z_j = z, Z_\ell = z') = \frac{N_{00}}{2^k} Q_0(z)Q_0(z') + \frac{N_{01}}{2^k} Q_0(z)Q_1(z') + \frac{N_{10}}{2^k} Q_1(z)Q_0(z') + \frac{N_{11}}{2^k} Q_1(z)Q_1(z') \tag{2.41}$$

$$= \frac{1}{4} Q_0(z)Q_0(z') + \frac{1}{4} Q_0(z)Q_1(z') + \frac{1}{4} Q_1(z)Q_0(z') + \frac{1}{4} Q_1(z)Q_1(z'). \tag{2.42}$$

Therefore,

$$\mathbb{E}_{\widehat{Q}^n}[T(\mathbf{Z})] = \frac{1}{|\mathcal{S}|} \sum_{j \in \mathcal{S}} \mathbb{E}_{\frac{1}{2}Q_0 + \frac{1}{2}Q_1} \left[\frac{Q_1(Z_j) - Q_0(Z_j)}{Q_0(Z_j)} \right] \tag{2.43}$$

$$= \sum_z \left(Q_0(z) + \frac{1}{2}(Q_1(z) - Q_0(z)) \right) \frac{Q_1(z) - Q_0(z)}{Q_0(z)} \tag{2.44}$$

$$= \frac{1}{2} \chi_2(Q_1 \| Q_0). \tag{2.45}$$

Similarly,

$$\mathbb{E}_{\widehat{Q}^n}\left[T(\mathbf{Z})^2\right] = \frac{1}{|\mathcal{S}|^2} \mathbb{E}_{\widehat{Q}^n} \left[\sum_{j \in \mathcal{S}} \frac{(Q_1(Z_j) - Q_0(Z_j))^2}{Q_0(Z_j)^2} + \sum_{j,\ell \in \mathcal{S}: j \neq \ell} \frac{Q_1(Z_j) - Q_0(Z_j)}{Q_0(Z_j)} \frac{Q_1(Z_\ell) - Q_0(Z_\ell)}{Q_0(Z_\ell)} \right] \tag{2.46}$$

$$= \frac{1}{|\mathcal{S}|} \chi_2(Q_1 \| Q_0) + \frac{1}{2\,|\mathcal{S}|} \chi_3(Q_1 \| Q_0) + \frac{1}{|\mathcal{S}|^2} \sum_{j,\ell \in \mathcal{S}: j \neq \ell} \mathbb{E}_{\widehat{Q}^n} \left[\frac{Q_1(Z_j) - Q_0(Z_j)}{Q_0(Z_j)} \frac{Q_1(Z_\ell) - Q_0(Z_\ell)}{Q_0(Z_\ell)} \right]. \tag{2.47}$$

Note that by (2.42), we have

$$\mathbb{E}_{\widehat{Q}^n}\left[\frac{Q_1(Z_j)-Q_0(Z_j)}{Q_0(Z_j)}\frac{Q_1(Z_\ell)-Q_0(Z_\ell)}{Q_0(Z_\ell)}\right]$$

$$= \mathbb{E}_{\frac{1}{4}Q_0Q_0+\frac{1}{4}Q_0Q_1+\frac{1}{4}Q_1Q_0+\frac{1}{4}Q_1Q_1}\left[\frac{Q_1(Z_j)-Q_0(Z_j)}{Q_0(Z_j)}\frac{Q_1(Z_\ell)-Q_0(Z_\ell)}{Q_0(Z_\ell)}\right] \tag{2.48}$$

$$= \frac{1}{4}\left(\mathbb{E}_{Q_1}\left[\frac{Q_1(Z)-Q_0(Z)}{Q_0(Z)}\right]\right)^2 = \frac{1}{4}\chi_2(Q_1\|Q_0)^2. \tag{2.49}$$

Therefore, combining (2.47) and (2.49), we obtain

$$\mathrm{Var}_{\widehat{Q}^n}\left(T(\mathbf{Z})\right) = \mathbb{E}_{\widehat{Q}^n}\left[T(\mathbf{Z})^2\right] - \mathbb{E}_{\widehat{Q}^n}[T(\mathbf{Z})]^2 \tag{2.50}$$

$$= \frac{1}{|\mathcal{S}|}\chi_2(Q_1\|Q_0) + \frac{1}{2\,|\mathcal{S}|}\chi_3(Q_1\|Q_0) + \frac{|\mathcal{S}|\,(|\mathcal{S}|-1)}{4\,|\mathcal{S}|^2}\chi_2(Q_1\|Q_0)^2 - \frac{1}{4}\chi_2(Q_1\|Q_0)^2 \tag{2.51}$$

$$= \frac{1}{|\mathcal{S}|}\chi_2(Q_1\|Q_0) + \frac{1}{2\,|\mathcal{S}|}\chi_3(Q_1\|Q_0) - \frac{1}{4\,|\mathcal{S}|}\chi_2(Q_1\|Q_0)^2. \tag{2.52}$$

Finally, to simplify the analysis, we choose a convenient threshold $\gamma = \frac{1}{4}\chi_2(Q_1\|Q_0)$. For the test $T(\mathbf{z})$ defined in (2.36) together with the threshold γ, the probability of false alarm α satisfies

$$\alpha \triangleq \mathbb{P}_{Q_0^{\otimes n}}(T(\mathbf{Z}) > \gamma) \leqslant \frac{\mathrm{Var}_{Q_0^{\otimes n}}\left(T(\mathbf{Z})\right)}{\gamma^2} = \frac{16}{|\mathcal{S}|\,\chi_2(Q_1\|Q_0)}. \tag{2.53}$$

Similarly,

$$\beta = \mathbb{P}_{\widehat{Q}^n}(T(\mathbf{Z}) < \gamma) = \mathbb{P}_{\widehat{Q}^n}(-T(\mathbf{Z}) > -\gamma) \tag{2.54}$$

$$= \mathbb{P}_{\widehat{Q}^n}\left(\mathbb{E}_{\widehat{Q}^n}[T(\mathbf{Z})] - T(\mathbf{Z}) > \frac{1}{2}\chi_2(Q_1\|Q_0) - \gamma\right) \tag{2.55}$$

$$\leqslant \mathbb{P}_{\widehat{Q}^n}\left(\left|\mathbb{E}_{\widehat{Q}^n}[T(\mathbf{Z})] - T(\mathbf{Z})\right| > \frac{1}{2}\chi_2(Q_1\|Q_0) - \gamma\right) \tag{2.56}$$

$$\leqslant \frac{\mathrm{Var}_{\widehat{Q}^n}\left(T(\mathbf{Z})\right)}{\gamma^2} \tag{2.57}$$

$$= \frac{16}{|\mathcal{S}|\,\chi_2(Q_1\|Q_0)} + \frac{8\chi_3(Q_1\|Q_0)}{|\mathcal{S}|\,\chi_2(Q_1\|Q_0)^2} - \frac{4}{|\mathcal{S}|}. \tag{2.58}$$

The result follows by recalling that $|\mathcal{S}| \geqslant k$ and $\beta \geqslant 0$. □

As a corollary of Proposition 4, we have the following.

Corollary 1. *A family of (n, k_n) binary linear codes with $(k_n, n) \in (\mathbb{N}^*)^2$ and $k_n \leqslant n$, for which $k_n = \omega(1)$ as n goes to infinity cannot achieve covert capacity for a BI-DMC without using shared secret key. In particular, linear codes cannot achieve any fraction of the covert capacity in this case.*

Proof. Consider a family of (n, k_n) codes with $k_n = \omega(1)$, i.e., $\lim_{n\to\infty} k_n = \infty$. By Proposition 4, there exists a test for which $\alpha = \mathcal{O}\left(\frac{1}{k_n}\right)$ and $\beta = \mathcal{O}\left(\frac{1}{k_n}\right)$ as n goes to infinity and the communication is detected with nonzero probability for n large enough. The result follows since achieving covert capacity would require $k_n = \theta\sqrt{n\delta}$ for $\delta \geqslant \mathbb{D}\left(\widehat{Q}^n \| Q_0^{\otimes n}\right)$ and some $\theta > 0$ [16]. □

CHAPTER 3

MULTILEVEL CODING FOR COVERT COMMUNICATIONS

In this chapter, we introduce our coding scheme for covert communication using multilevel coding with multistage decoding and provide an information-theoretic analysis of the scheme. As mentioned in the previous chapter, most of the research on covert communication has focused on the information-theoretic analysis, which proves the existence of codes using random coding argument. There has been little attention paid to the design of codes for covert communication with an explicit construction that can be implemented in practice. In this chapter, we propose a coding scheme that achieves the covert capacity for BI-DMCs with a focus on practical implementation. One of the important points to note while designing codes for covert communication is that the codeword weights should be of the order of $\sqrt{n}$. There is a conjecture that this requirement may not be achieved using a binary linear coding without introducing any non-linearity to the scheme. In fact, we know from Section 2.6 that it is impossible to achieve covert capacity using a binary linear code without any secret key shared between the legitimate users. In theory, a shared secret is not required when the Willie's channel is degraded with respect to (w.r.t.) Bob's channel. However, even for this degraded case, linear codes cannot be optimal for covert communication.

Motivated by this negative result, we have developed a non-linear scheme in which the non-linearity is introduced through PPM. The results in [44] show that PPM achieves the covert capacity when combined with suitable random non-binary codes. One subtlety behind the results in [44] is that the PPM order (and therefore the non-binary field size) grows *linearly* with the blocklength ℓ of the code used over the super channel as

$$\ell = \left\lceil \frac{2\delta}{\chi_2(Q_1 \| Q_0)} m \right\rceil, \tag{3.1}$$

making it hard to realize such codes in practice with low-complexity. Reed-Solomon (RS) codes offer the desired scaling (they are q-ary code of length $q-1$), but not the flexibility needed to meet the requirement in (3.1). To circumvent the design of non-binary codes, we use MLC to decompose the PPM super-channel into $q \in \mathbb{N}^*$ binary-input channels.

3.1 Notation

In this section, we introduce the main notations that we use in this chapter and subsequent ones. Let $[\![a,b]\!]$ be the set of integers between $\lfloor a \rfloor$ and $\lceil b \rceil$. When used as a subscript or a superscript, we denote $[\![a,b]\!]$ by $a{:}b$. As we use PPM extensively in this chapter, we recall the definition of PPM symbol and define the related notations. For $q \in \mathbb{N},\ m = 2^q$, and $i \in [\![1,m]\!]$, we define a PPM symbol $\widetilde{x}_i$ of order m as the binary vector of length m such that the i-th component is one and all other components are zero. The use of PPM modulation over the channel BI-DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ defines a "super channel" with transition probability $\widetilde{W}_{\widetilde{Y}|\widetilde{X}} \triangleq W_{Y|X}^{\otimes m}$, whose input alphabet is the set of all PPM symbols of order m denoted by $\widetilde{\mathcal{X}}_q = \{\widetilde{x}_i\}_{i=1}^m$ and whose output alphabet is $\widetilde{\mathcal{Y}}_q = \mathcal{Y}^{2^q}$. For any set $\mathcal{S} \subseteq [\![1,q]\!]$ and a sequence of random variables $X_1, \ldots, X_q$, we denote the set of random variables $(X_i)_{i\in\mathcal{S}}$ by $X_{\mathcal{S}}$. Using this notation, we have $X_{1:q} = (X_1, \ldots, X_q)$. For a binary sequence $x_{1:q} \in \mathbb{F}_2^q$, we let $d(x_{1:q}) \in \mathbb{N}$ denote the decimal equivalent of $x_{1:q}$ defined as

$$d(x_{1:q}) \triangleq 1 + \sum_{i=1}^{q} x_i \cdot 2^{i-1}.$$

Let $d^{-1}(\cdot)$ denote the inverse operation of $d(\cdot)$, that is, $d^{-1}(d(x_{1:q})) = x_{1:q}$. For $j \in [\![1,q]\!]$, we call $d^{-1}(j)$ the binary equivalent of j. Note that the operation $d^{-1}(\cdot)$ is specific to the value of q even though it is not explicitly evident in the notation. Let $\mathcal{A}^q(x_{\mathcal{S}})$ denote the elements of $[\![1,2^q]\!]$ that agree in their binary equivalent with $x_{1:q}$ in positions $\mathcal{S}$; that is, for

any set $\mathcal{S} \subseteq [\![1, q]\!]$ and $x_{1:q} \in \mathbb{F}_2^q$, we define

$$\mathcal{A}^q(x_{\mathcal{S}}) \triangleq \left\{ j \in [\![1, 2^q]\!] : (d^{-1}(j))_{\mathcal{S}} = x_{\mathcal{S}} \right\}, \tag{3.2}$$

where $(d^{-1}(\cdot))_{\mathcal{S}}$ denotes the components of $d^{-1}(\cdot)$ indexed by the elements of $\mathcal{S}$. The complement of the set $\mathcal{A}^q(x_{\mathcal{S}})$ w.r.t. $[\![1, 2^q]\!]$ is denoted by $\mathcal{A}^q(x_{\mathcal{S}})^c \triangleq [\![1, 2^q]\!] \backslash \mathcal{A}^q(x_{\mathcal{S}})$. For example, let $q = 4$, $x_{1:q} = 1011$, and $\mathcal{S} = \{2, 3\}$, then $\mathcal{A}^q(x_{\mathcal{S}}) = \{5, 6, 13, 14\}$.

We define a PPM mapper as a one-to-one mapping from the set of all binary vectors of length q to the set of all PPM symbols of order 2^q denoted by

$$\widetilde{x} : \mathbb{F}_2^q \to \widetilde{\mathcal{X}}_q : x_{1:q} \mapsto \widetilde{x}(x_{1:q}) = \widetilde{x}_{\mathcal{A}^q(x_{1:q})} = \widetilde{x}_{d(x_{1:q})}, \tag{3.3}$$

Hence, we denote the PPM super channel equivalently in terms of the binary vector input as $\widetilde{W}_{\widetilde{Y}|\widetilde{X}} \equiv \widetilde{W}_{\widetilde{Y}|X_{1:q}}$. We use both notations interchangeably depending on the situation. For the channel $\widetilde{W}_{\widetilde{Y}|\widetilde{X}}$, the distribution at the output by a uniformly distributed PPM symbols of order m as the input is denoted by P^m_{PPM} and is given by

$$P^m_{\text{PPM}}(\widetilde{y}) \triangleq \frac{1}{m} \sum_{i=1}^{m} \widetilde{W}_{\widetilde{Y}|\widetilde{X}}(\widetilde{y}|\widetilde{x}_i) = \frac{1}{2^q} \sum_{x_{1:q} \in \mathcal{X}^q} \widetilde{W}_{\widetilde{Y}|X_{1:q}}(\widetilde{y}|x_{1:q}). \tag{3.4}$$

Recall from Section 2.3 that we use BI-DMC $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ as the main channel or Bob's channel and BI-DMC $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$ as the warden's channel or Willie's channel. The distribution induced at the output of Willie's channel by a uniformly distributed PPM symbols is denoted by $Q^m_{\text{PPM}}(\widetilde{y})$ and is defined similar to (3.4) with $W_{Z|X}$ instead of $W_{Y|X}$.

3.2 Setup for MLC with PPM

Alice divides her uniform message W into q independent and uniformly distributed messages $(W_1, \ldots, W_q)$ where W_i represents the input message to the i-th level, which we will precisely define later. The setup is illustrated in Fig. 3.1, in which q binary encoders feed

their codeword outputs to a PPM symbol mapper of order $m \triangleq 2^q$. In the i-th level, the encoder encodes message W_i into a binary vector $\mathbf{X}_i \in \mathbb{F}_2^\ell$ of length ℓ. The PPM mapper takes these coded binary vectors from each level in parallel and maps the binary vector of length q entering the mapper to a PPM symbol; we thus obtain a vector of PPM symbols, $\widetilde{\mathbf{X}} \in \widetilde{\mathcal{X}}_q^\ell$ at the output of the PPM mapper. Bob and Willie observe the outputs of channels $\widetilde{W}_{\widetilde{Y}|\widetilde{X}}$ and $\widetilde{W}_{\widetilde{Z}|\widetilde{X}}$ denoted by $\widetilde{\mathbf{Y}} \in \widetilde{\mathcal{Y}}_q^\ell$ and $\widetilde{\mathbf{Z}} \in \widetilde{\mathcal{Z}}_q^\ell$, respectively. Let $Q_0^{\otimes m\ell}$ denote the innocent distribution induced when only innocent symbol "0" is sent through the channel $W_{Z|X}$ over $m\ell$ channel uses, and $P_{\widetilde{\mathbf{Z}}}$ denote the distribution induced at Willie's receiver by coding over ℓ super-channel uses. Before analyzing the coding scheme further, it is worth noting that the benefits of MLC are not a priori obvious. In fact, since the number of levels changes with the order of the PPM symbol m and the blocklength ℓ, which are related by (3.1), one could expect that the channels perceived at each level $i \in [\![1, q]\!]$ would vary as the blocklength grows, making code design challenging. Perhaps surprisingly, we show in Chapter 4 that this is *not* the case and that codes may be designed for fixed channels; this is particularly convenient as it allows us to exploit families of channel capacity- and channel resolvability-achieving codes, such as polar codes [46, 38].

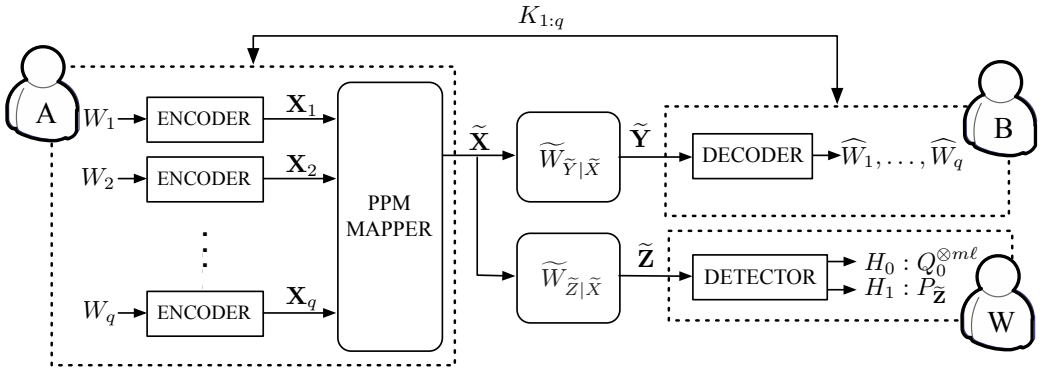

Figure 3.1: Setup for MLC over PPM super channel.

3.3 Information-theoretic analysis of MLC

We now present the information-theoretic analysis of MLC and show using a random coding argument that there exist binary coding schemes with MLC that achieve the covert capacity of BI-DMCs. From [17, Theorem 2], we know that we need a key to achieve covertness only when $\mathbb{D}(P_1||P_0) < \mathbb{D}(Q_1||Q_0)$. However, in MLC, we are coding over multiple levels and the channel corresponding to individual levels may not preserve the same relation. Consequently, irrespective of the relationship between the original channels, we might need keys for some levels, whereas we might be able to send some covert and secret messages over some other levels depending on the mutual information relations between the two channels corresponding to that particular level. We can overcome this by using chaining, which consists of breaking the transmission into several blocks and using the secret messages from one block as the keys for the next block. By using chaining, we only require extra keys for the first block, and on average the throughputs of messages and keys tend to the optimal throughputs as the number of blocks tends to infinity. We prove this by first characterizing the size of keys and messages for a single block and then developing a chaining strategy that achieves optimal throughput and covertness over B blocks.

We first describe the coding scheme for a single block. Let $W_i = (U_i, V_i)$ with $U_i \in [\![1, M_{U,i}]\!]$ and $V_i \in [\![1, M_{V,i}]\!]$ be a random variable that represents the message for the i-th level and $K_i \in [\![1, M_{K,i}]\!]$ be the key shared between Alice and Bob. V_i represents the secret part of the message, which is reused as a key for the next block. U_i, V_i, and K_i are independent and uniformly distributed random variables. Let $R_{U,i} \triangleq \log M_{U,i}/\ell$, $R_{V,i} \triangleq \log M_{V,i}/\ell$, and $R_{K,i} \triangleq \log M_{K,i}/\ell$ represent the rates of U_i, V_i, and K_i, respectively. Encoders for each level are independent; therefore, we can view the MLC scheme in Fig.3.1 as a q-user MAC in which the PPM mapper is integrated to the channel. Because of the symmetry of the PPM super channel, a uniform distribution on PPM symbols achieves the capacity for Bob's channel. Note that, to ensure (3.1) holds, the number of levels q depends

on the blocklength ℓ of the code. Hence, we cannot directly use the results for MAC with a fixed number of users [47]. After careful analysis, we show that the result remains what would have expected from [47].

The codebook, encoding, and decoding for one block are as follows.

- *Codebook*: For i-th level, generate a codebook of $M_{U,i} \times M_{V,i} \times M_{K,i}$ codewords, each of length ℓ, denoted by $\mathbf{X}_i(u_i, v_i, k_i)$, where $u_i \in [\![1, M_{U,i}]\!]$, $v_i \in [\![1, M_{V,i}]\!]$, and $k_i \in [\![1, M_{K,i}]\!]$. Each codeword is generated independently according to a uniform distribution on $\mathcal{X}^\ell$, denoted by $P_{X_i}^{\otimes \ell}$.

- *Encoding*: Given messages U_i and V_i and key K_i for the i-th level, the i-th encoder selects the codeword $\mathbf{X}_i(U_i, V_i, K_i)$.

- *Decoding*: The receiver decodes $\hat{\mathbf{w}} = (\hat{u}_1, \ldots, \hat{u}_q, \hat{v}_1, \ldots, \hat{v}_q)$ from $\widetilde{\mathbf{y}}$ knowing the key $\mathbf{k} = (k_1, \ldots, k_q)$ if $\hat{\mathbf{w}}$ is the unique message such that $(\mathbf{X}_1(\hat{u}_1, \hat{v}_1, k_1), \ldots, \mathbf{X}_q(\hat{u}_q, \hat{v}_q, k_q), \widetilde{\mathbf{y}}) \in \mathcal{T}_\Gamma^\ell$, where $\mathcal{T}_\Gamma^\ell$ is defined as follows. For $\Gamma = \{\gamma_\mathcal{S}\}_{\mathcal{S} \subseteq [\![1,q]\!]}$,

$$\mathcal{T}_\Gamma^\ell \triangleq \left\{ (\mathbf{x}_{1:q}, \widetilde{\mathbf{y}}) \in \mathcal{X}^{q\ell} \times \widetilde{\mathcal{Y}}^\ell : \log \frac{W_{\widetilde{Y}|X_{1:q}}^{\otimes \ell}(\widetilde{\mathbf{y}}|\mathbf{x}_{1:q})}{P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{S}^c})} > \gamma_\mathcal{S}, \quad \forall \mathcal{S} \subseteq [\![1, q]\!] \right\}, \quad (3.5)$$

where $P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}$ is the conditional distribution given by

$$P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{S}^c}) = \sum_{\mathbf{x}_\mathcal{S}} \prod_{j \in \mathcal{S}} P_{X_j}^{\otimes \ell}(\mathbf{x}_j) W_{\widetilde{Y}|X_{1:q}}^{\otimes \ell}(\widetilde{\mathbf{y}}|\mathbf{x}_{1:q}) \quad = \prod_{i=1}^{\ell} P_{\widetilde{Y}|X_{\mathcal{S}^c}}(\widetilde{y}_i|x_{\mathcal{S}^c,i}) \,. \quad (3.6)$$

To simplify the notation, we use the following definitions:

$$\mathbf{u} \triangleq (u_1, \ldots, u_q), \ \mathbf{v} \triangleq (v_1, \ldots, v_q), \ \mathbf{w} \triangleq (\mathbf{u}, \mathbf{v}),$$

$$\mathbf{X}_{1:q}(\mathbf{w}, \mathbf{k}) \triangleq \mathbf{X}_{1:q}(\mathbf{u}, \mathbf{v}, \mathbf{k}) \triangleq (\mathbf{X}_1(u_1, v_1, k_1), \ldots, \mathbf{X}_q(u_q, v_q, k_q)),$$

$$X_{1:q,i}(\mathbf{w}, \mathbf{k}) \triangleq X_{1:q,i}(\mathbf{u}, \mathbf{v}, \mathbf{k}) \triangleq (X_{1,i}(u_1, v_1, k_1), \ldots, X_{q,i}(u_q, v_q, k_q),$$

$$\mathbf{X}_\mathcal{G}(\mathbf{w}, \mathbf{k}) \triangleq (\mathbf{X}_i(u_i, v_i, k_i))_{i \in \mathcal{G}}, \quad \forall \mathcal{G} \subseteq [\![1, q]\!]; \qquad M_U \triangleq M_{U,1} \times \cdots \times M_{U,q},$$

$$M_V \triangleq M_{V,1} \times \cdots \times M_{V,q}, \quad M_K \triangleq M_{K,1} \times \cdots \times M_{K,q}, \quad M \triangleq M_U \times M_V.$$

We first analyze channel reliability and channel resolvability for one block and prove the existence of a random code that satisfies the requirements for covert communication. Then, we use chaining over blocks using the same code for each block to obtain a coding scheme that achieves optimal covert throughputs. For chaining over B blocks, the variables in the j-th block is identified by a superscript; for example, we denote the message vector by $\mathbf{w}^{(j)}$, the key vector by $\mathbf{k}^{(j)}$, the corresponding codeword for the i-th level by $\mathbf{x}_i(\mathbf{w}^{(j)}, \mathbf{k}^{(j)}) \triangleq \mathbf{x}_i(w_i^{(j)}, k_i^{(j)})$, and the outputs of Bob's channel and Willie's channel by $\widetilde{\mathbf{y}}^{(j)}$ and $\widetilde{\mathbf{z}}^{(j)}$ respectively.

3.3.1 Reliability analysis

We now analyze the rate requirements for achieving vanishing probability of error using random codes for a single block. We denote the probability of error for a given codebook by $\mathbb{P}\left(\widehat{\mathbf{W}} \neq \mathbf{W}\right)$ and the expectation of it over the random codebook distribution by $\mathbb{E}\left[\mathbb{P}\left(\widehat{\mathbf{W}} \neq \mathbf{W}\right)\right]$. The following lemma summarizes the rate requirement.

Lemma 3. *For $\mathcal{S} \subseteq [\![1, q]\!]$ and $0 < \epsilon < \varepsilon/2$, if the rates of random codes satisfy*

$$\sum_{i \in \mathcal{S}} (R_{U,i} + R_{V,i}) \leqslant I(X_{\mathcal{S}}; \widetilde{Y} | X_{\mathcal{S}^c}) - \frac{\varepsilon}{q}, \tag{3.7}$$

then

$$\mathbb{E}\left[\mathbb{P}\left(\widehat{\mathbf{W}} \neq \mathbf{W}\right)\right] \leqslant \delta_1 \triangleq m \left(e^{-\frac{\ell \epsilon^2}{2(\log 2)^2 q^4}} + e^{-\frac{\ell \epsilon}{q}} \right). \tag{3.8}$$

Proof. See Appendix 3.A. □

3.3.2 Covertness analysis

We now analyze the channel resolvability of Willie's channel and prove the existence of channel resolvability codes if the rates satisfy some condition. Let $P_{\widetilde{\mathbf{Z}}}$ be the distribution induced by the code for a given block, Q_{PPM}^m be the output distribution of Willie's channel when the input is uniform over all possible PPM symbols of order m, and $Q_0^{\otimes m\ell}$ be the innocent distribution. By calculations similar to [17, (77)-(79)], we have

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|Q_0^{\otimes m\ell}\big) \leqslant \mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|(Q_{\text{PPM}}^m)^{\otimes\ell}\big) + 2\ell \times \mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes\ell}\big) \max_{\widetilde{z}} \left|\log \frac{(Q_{\text{PPM}}^m)(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})}\right| + \mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes\ell}\|Q_0^{\otimes m\ell}\big). \quad (3.9)$$

From Pinsker's inequality, we have

$$\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes\ell}\big) \leqslant \sqrt{\frac{1}{2}\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|(Q_{\text{PPM}}^m)^{\otimes\ell}\big)}. \quad (3.10)$$

We show using the following lemma that the first two terms go to zero exponentially in ℓ.

Lemma 4. *For $\mathcal{S} \subseteq [\![1, q]\!]$ and $0 < \epsilon < \varepsilon/2$, if the rates of random codes satisfy*

$$\sum_{i\in\mathcal{S}}(R_{U,i} + R_{K,i}) \geqslant I(X_{\mathcal{S}}; \widetilde{Z}) + \frac{\varepsilon}{q}, \quad (3.11)$$

then

$$\mathbb{E}\left[\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}|\mathbf{V}=\mathbf{v}}\|\left(Q_{\text{PPM}}^m\right)^{\otimes\ell}\Big)\right] \leqslant \delta_2 \triangleq me^{-\ell\epsilon/q} + \frac{m^2\ell 2^{q/m\ell}}{\mu_z} e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}}, \quad (3.12)$$

where $\mu_z = \min_{z\in\text{supp}(Q_Z)} Q_Z(z)$ with $Q_Z(z) = \frac{1}{2}\sum_{x\in\mathcal{X}} W_{Z|X}(z|x)$.

Proof. See Appendix 3.B. □

Lemma 4 actually proves a stronger result than what is required to bound (3.9), as it guarantees that, on average over all possible values of the message $\mathbf{v}$, the distribution

induced by coding remains identical regardless of the values of $\mathbf{v}$. This stronger result will prove useful in Section 3.3.5 when we integrate the code in a chained construction.

3.3.3 Identifying a specific code

Note that Lemma 3 and Lemma 4 are obtained by taking expectation over randomly generated sequence of codes. We now show that there exists a specific sequence of codes that satisfies the reliability and covertness conditions. We denote the probability measure w.r.t. codebook distribution by $\mathbb{P}_{\mathcal{C}}(\cdot)$. Using Markov's inequality, for $\eta_1 > 0$ and $\eta_2 > 0$, we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}}\left(\mathbb{P}\Big(\widehat{\mathbf{W}} \neq \mathbf{W}\Big) < \eta_1\delta_1, \sum_{\mathbf{v}} \frac{1}{M_V}\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\mathbf{V}^{(j)}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) < \eta_2\delta_2\right) \\
&\quad = 1 - \mathbb{P}_{\mathcal{C}}\left(\mathbb{P}\Big(\widehat{\mathbf{W}} \neq \mathbf{W}\Big) \geqslant \eta_1\delta_1 \text{ or } \sum_{\mathbf{v}} \frac{1}{M_V}\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\mathbf{V}^{(j)}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) \geqslant \eta_2\delta_2\right) \\
&\quad \geqslant 1 - \mathbb{P}_{\mathcal{C}}\Big(\mathbb{P}\Big(\widehat{\mathbf{W}} \neq \mathbf{W}\Big) \geqslant \eta_1\delta_1\Big) \\
&\qquad\qquad - \mathbb{P}_{\mathcal{C}}\left(\sum_{\mathbf{v}} \frac{1}{M_V}\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\mathbf{V}^{(j)}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) \geqslant \eta_2\delta_2\right) \qquad (3.13) \\
&\quad \geqslant 1 - \frac{1}{\eta_1} - \frac{1}{\eta_2}. \qquad (3.14)
\end{aligned}$$

By choosing η_1 and η_2 such that $\frac{1}{\eta_1} + \frac{1}{\eta_2} < 1$, we have the above probability positive, which means there exist codes that satisfy

$$\mathbb{P}\Big(\widehat{\mathbf{W}} \neq \mathbf{W}\Big) < \eta_1\delta_1, \qquad (3.15)$$

$$\sum_{\mathbf{v}} \frac{1}{M_V}\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\mathbf{V}^{(j)}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) < \eta_2\delta_2. \qquad (3.16)$$

Moreover, because of the convexity of divergence, we have

$$\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) \leqslant \sum_{\mathbf{v}} \frac{1}{M_V}\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\mathbf{V}^{(j)}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) < \eta_2\delta_2. \qquad (3.17)$$

Since δ_2 goes to zero exponentially in ℓ, the first two terms of (3.9) also go to zero exponentially in ℓ. Using [44, Lemma 1], we have

$$\mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes \ell} \| Q_0^{\otimes m\ell}\big) \leqslant \frac{\ell}{m}\left(\frac{\chi_2(Q_1\|Q_0)}{2} + \mathcal{O}(\frac{1}{m})\right). \tag{3.18}$$

By choosing $\ell = \lfloor \frac{2m\delta}{B\chi_2(Q_1\|Q_0)} \rfloor$,

$$\mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes \ell} \| Q_0^{\otimes m\ell}\big) \leqslant \frac{\delta}{B} + \mathcal{O}(\frac{1}{Bm}). \tag{3.19}$$

Hence, we have

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| Q_0^{\otimes m\ell}\big) \leqslant \frac{\delta}{B} + \mathcal{O}(\frac{1}{Bm}). \tag{3.20}$$

From (3.16), we also obtain the secrecy of $\mathbf{V}$ as follows.

$$\begin{aligned} I\left(\widetilde{\mathbf{Z}}; \mathbf{V}\right) &= \mathbb{D}\big(P_{\widetilde{\mathbf{Z}}\mathbf{V}} \| P_{\widetilde{\mathbf{Z}}} P_{\mathbf{V}}\big) && (3.21)\\ &= \sum_{\mathbf{v}} \frac{1}{M_V} \mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}|\mathbf{V}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) - \mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) && (3.22)\\ &\leqslant \sum_{\mathbf{v}} \frac{1}{M_V} \mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}|\mathbf{V}=\mathbf{v}} \| \left(Q_{\text{PPM}}^m\right)^{\otimes \ell}\Big) && (3.23)\\ &\leqslant \eta_2 \delta_2. && (3.24) \end{aligned}$$

3.3.4 MSD operating point for MLC

We know from [44] that the capacity of the PPM super channel converges to $\mathbb{D}(P_1\|P_0)$ as the PPM order m tends to infinity. Since the PPM mapper is a one-to-one map between $X_{1:q}$ and the PPM symbol $\widetilde{X} = \widetilde{x}(X_{1:q})$, we have the following decomposition of mutual

information between the input and the output of the PPM super channel $\widetilde{W}_{\widetilde{Y}|\widetilde{X}}$.

$$I(\widetilde{X};\widetilde{Y}) = I(X_{1:q};\widetilde{Y}) = \sum_{i=1}^{q} I(X_i;\widetilde{Y}|X_{i+1:q}). \tag{3.25}$$

The term $I(X_i;\widetilde{Y}|X_{i+1:q})$ represents the average mutual information between the input of the i-th level and $\widetilde{Y}$ for a given input to the levels $i+1$ to q. Since X_i is independent of $X_{i+1:q}$, we have

$$I(X_i;\widetilde{Y}|X_{i+1:q}) = I(X_i;\widetilde{Y},X_{i+1:q}). \tag{3.26}$$

This suggests that we can consider the i-th level as a channel with input X_i and output $(\widetilde{Y},X_{i+1:q})$ and decode all levels successively in descending order from q to 1 using MSD; specifically, we decode the q-th level first, then the $(q-1)$-th level by treating the decoded bits from q-th level as a side information for the channel defining $(q-1)$-th level, and so on. From [48], we know that the capacity of the PPM super channel could be achieved using MSD by choosing the corresponding terms in the summation of (3.25) as the rates for each level. In our problem, we show that by choosing the rates for i-th level as follows, we can achieve the optimal rate.

$$R_{U,i} = \min\left(I(X_i;\widetilde{Y}|X_{i+1:q}) - \varepsilon/q, I(X_i;\widetilde{Z}|X_{i+1:q}) + \varepsilon/q\right), \tag{3.27}$$

$$R_{V,i} = \max\left(0, I(X_i;\widetilde{Y}|X_{i+1:q}) - I(X_i;\widetilde{Z}|X_{i+1:q}) - 2\varepsilon/q\right), \tag{3.28}$$

$$R_{K,i} = \max\left(0, I(X_i;\widetilde{Z}|X_{i+1:q}) - I(X_i;\widetilde{Y}|X_{i+1:q}) + 2\varepsilon/q\right). \tag{3.29}$$

The sum in (3.25) converges to $\mathbb{D}(P_1\|P_0)$ as q tends to infinity. This suggests that we can achieve rates arbitrarily close to $\mathbb{D}(P_1\|P_0)$ by choosing the number of levels in MLC large enough and rates given in (3.27-3.29) for each level.

We now show that the rates in (3.27-3.29) satisfy the rate requirements in (3.7) and

(3.11). For $\mathcal{S} \subseteq [\![1, q]\!]$, we have

$$\sum_{i\in\mathcal{S}}(R_{U,i} + R_{V,i}) = \sum_{i\in\mathcal{S}} \Big(I(X_i; \widetilde{Y}|X_{i+1:q}) - \varepsilon/q\Big) \tag{3.30}$$

$$= \sum_{i\in\mathcal{S}} I(X_i; \widetilde{Y}|X_{[\![i+1,q]\!]\cap\mathcal{S}}, X_{[\![i+1,q]\!]\cap\mathcal{S}^c}) - |\mathcal{S}|\varepsilon/q \tag{3.31}$$

$$= \sum_{i\in\mathcal{S}} \Big(H(X_i|X_{[\![i+1,q]\!]\cap\mathcal{S}}, X_{[\![i+1,q]\!]\cap\mathcal{S}^c}) - H(X_i|\widetilde{Y}, X_{[\![i+1,q]\!]\cap\mathcal{S}}, X_{[\![i+1,q]\!]\cap\mathcal{S}^c})\Big) - |\mathcal{S}|\varepsilon/q \tag{3.32}$$

$$\overset{(a)}{\leqslant} \sum_{i\in\mathcal{S}} \Big(H(X_i|X_{[\![i+1,q]\!]\cap\mathcal{S}}, X_{\mathcal{S}^c}) - H(X_i|\widetilde{Y}, X_{[\![i+1,q]\!]\cap\mathcal{S}}, X_{\mathcal{S}^c})\Big) - |\mathcal{S}|\varepsilon/q \tag{3.33}$$

$$= I(X_{\mathcal{S}}, \widetilde{Y}|X_{\mathcal{S}^c}) - |\mathcal{S}|\varepsilon/q, \tag{3.34}$$

where (a) follows from the fact that conditioning reduces entropy and the X_i's are independent of each other. Hence, the rates satisfy the reliability rate constraints in (3.7). Similarly,

$$\sum_{i\in\mathcal{S}}(R_{U,i} + R_{K,i}) = \sum_{i\in\mathcal{S}} \Big(I(X_i; \widetilde{Z}|X_{i+1:q}) + \varepsilon/q\Big) \tag{3.35}$$

$$= \sum_{i\in\mathcal{S}} I(X_i; \widetilde{Z}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}, \{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}) + |\mathcal{S}|\varepsilon/q \tag{3.36}$$

$$= \sum_{i\in\mathcal{S}} I(X_i; \widetilde{Z}, \{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) - I(X_i; \{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) + |\mathcal{S}|\varepsilon/q \tag{3.37}$$

$$\overset{(a)}{=} \sum_{i\in\mathcal{S}} I(X_i; \widetilde{Z}, \{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) + |\mathcal{S}|\varepsilon/q \tag{3.38}$$

$$\geqslant \sum_{i\in\mathcal{S}} I(X_i; \widetilde{Z}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) + |\mathcal{S}|\varepsilon/q \tag{3.39}$$

$$= I(X_{\mathcal{S}}; \widetilde{Z}) + |\mathcal{S}|\varepsilon/q, \tag{3.40}$$

where (a) follows from the independence of $\{X_i\}$. This shows that the rates satisfy the resolvability rate constraints in (3.11).

We now compute the covert message and key throughputs. The number of bits transmitted is given by

$$\log M_U + \log M_V = \ell \sum_{i=1}^{q} (R_{U,i} + R_{V,i}) \tag{3.41}$$

$$= \ell \left(I(X_{1:q}; \widetilde{Y}) - \varepsilon \right) \tag{3.42}$$

$$= \ell \left(I(\widetilde{X}; \widetilde{Y}) - \varepsilon \right) \tag{3.43}$$

$$\geqslant \ell \left(\mathbb{D}(P_1 \| P_0) - \frac{1}{m} \chi_2(P_1 \| P_0) - \varepsilon \right), \tag{3.44}$$

where the last inequality follows from [44, Lemma 2]. The covert message throughput is given by

$$\frac{\log M_U + \log M_V}{\sqrt{\ell m \delta / B}} \geqslant \sqrt{\frac{B\ell}{m\delta}} \left(\mathbb{D}(P_1 \| P_0) - \frac{1}{m} \chi_2(P_1 \| P_0) - \varepsilon \right) \tag{3.45}$$

$$= \sqrt{\frac{2}{\chi_2(Q_1 \| Q_0)}} \left(\mathbb{D}(P_1 \| P_0) - \frac{1}{m} \chi_2(P_1 \| P_0) - \varepsilon \right). \tag{3.46}$$

The number of key bits used is given by

$$\log M_K = \ell \sum_{i=1}^{q} R_{K,i} \tag{3.47}$$

$$\leqslant \ell \sum_{i=1}^{q} \max \left(0, I(X_i; \widetilde{Z} | X_{i+1:q}) - I(X_i; \widetilde{Y} | X_{i+1:q}) + 2\varepsilon/q \right), \tag{3.48}$$

and the covert key throughput is given by

$$\frac{\log M_K}{\sqrt{\ell m \delta / B}} \leqslant \sqrt{\frac{2}{\chi_2(Q_1 \| Q_0)}} \left(\sum_{i=1}^{q} \max \left(0, I(X_i; \widetilde{Z} | X_{i+1:q}) - I(X_i; \widetilde{Y} | X_{i+1:q}) + 2\varepsilon/q \right) \right). \tag{3.49}$$

We now show using an example that the above key rate is not optimal. Let Bob's channel be a BSC with probability of flipping $P_0(1) = P_1(0) = 0.2$ and Willie's channel be

a binary asymmetric channel (BAC) with probabilities of flipping $Q_0(1) = 0.1$ and $Q_1(0) = 0.4$. The relative entropies for these channels are $\mathbb{D}(P_1\|P_0) = 1.2$ and $\mathbb{D}(Q_1\|Q_0) = 1.083$. For this case, ideally, we would not need any key to achieve covert communication; however, from the computation of $I(X_i; \widetilde{Z}|X_{i+1:q}) - I(X_i; \widetilde{Y}|X_{i+1:q})$ for $q = 9$ levels shown in Table 3.1, we can conclude that we need some key bits since $I(X_i; \widetilde{Z}|X_{i+1:q}) - I(X_i; \widetilde{Y}|X_{i+1:q}) > 0$ for levels 4 to 9.

Table 3.1: $I(X_i; \widetilde{Z}|X_{i+1:q}) - I(X_i; \widetilde{Y}|X_{i+1:q})$ computed in *nats* for 9 levels.

1	2	3	4	5	6	7	8	9
-0.0627	-0.0313	-0.0058	**0.0054**	**0.0058**	**0.0036**	**0.0019**	**0.0010**	**0.0005**

Note that $I(X_i; \widetilde{Z}|X_{i+1:q}) - I(X_i; \widetilde{Y}|X_{i+1:q})$ is negative for the first 3 levels. This suggests that we can send some secret bits over the first 3 levels as $R_{V,i}$ is positive for those levels. Moreover, the absolute value of the sum of negative terms is greater than the sum of positive terms. This means that we can *generate* more keys than we consume. Furthermore, we can use chaining so that the secret messages from one block can be used as the keys for the next block.

3.3.5 Chaining over B blocks

We now show that by modifying the coding scheme by chaining over B blocks such that the secret messages from i-th block $\{V_j^{(i)}\}_{j\in[\![1,q]\!]}$ are used as the keys for $(i+1)$-th block $\{K_j^{(i+1)}\}_{j\in[\![1,q]\!]}$, we can achieve the optimal throughputs. We can aggregate the secret messages from the levels that support secret messages and distribute them across the levels for which we need a key. We first show that the probability of error goes to zero asymptotically. To bound the probability of error for B blocks with chaining, define $\mathcal{E}^{(i)} = \{\widehat{\mathbf{W}}^{(i)} \neq \mathbf{W}^{(i)}\}$ and $\mathcal{E}^{(1:B)} = \bigcup_{i=1}^{B} \mathcal{E}^{(i)}$, where $\mathbf{W}^{(i)}$ is the transmitted message and $\widehat{\mathbf{W}}^{(i)}$ is the decoded message at the receiver for the i-th block. Following steps similar to those in the proof of

[42, Lemma 8], we obtain

$$\mathbb{P}\big(\mathcal{E}^{(1:B)}\big) \leqslant \sum_{i=1}^{B} \mathbb{P}\Big(\mathcal{E}^{(i)} \,\Big|\, \mathcal{E}^{(i-1)^c}\Big) \tag{3.50}$$

$$\leqslant Bm\left(e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}} + e^{-\frac{\ell\epsilon}{q}}\right). \tag{3.51}$$

This shows that the probability of error goes to zero asymptotically.

We now show that the relative entropy between the induced and innocent distributions is upper bounded by δ asymptotically. Let $P_{\widetilde{\mathbf{Z}}^{(1:B)}}$ be the distribution induced by the coding scheme for B blocks. We want $P_{\widetilde{\mathbf{Z}}^{(1:B)}}$ to be close to the innocent distribution $Q_0^{\otimes Bm\ell}$. We have

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}^{(1:B)}} \| Q_0^{\otimes Bm\ell}\big) = \sum_{j=1}^{B} \mathbb{E}_{\widetilde{\mathbf{Z}}^{(j+1:B)}}\left[\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\widetilde{\mathbf{Z}}^{(j+1:B)}} \| Q_0^{\otimes m\ell}\Big)\right] \tag{3.52}$$

$$= \sum_{j=1}^{B} \Big(\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}^{(j)}} \| Q_0^{\otimes m\ell}\big) + \mathbb{E}_{\widetilde{\mathbf{Z}}^{(j+1:B)}}\left[\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}|\widetilde{\mathbf{Z}}^{(j+1:B)}} \| P_{\widetilde{\mathbf{Z}}^{(j)}}\Big)\right]\Big) \tag{3.53}$$

$$= \sum_{j=1}^{B} \left[\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}^{(j)}} \| Q_0^{\otimes m\ell}\big) + I(\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{Z}}^{(j+1:B)})\right], \tag{3.54}$$

and

$$I(\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{Z}}^{(j+1:B)}) \leqslant I(\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{Z}}^{(j+1:B)}, \mathbf{V}^{(j)}) \tag{3.55}$$

$$= I(\widetilde{\mathbf{Z}}^{(j)}; \mathbf{V}^{(j)}) + I(\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{Z}}^{(j+1:B)} | \mathbf{V}^{(j)}) \tag{3.56}$$

$$\overset{(a)}{=} I(\widetilde{\mathbf{Z}}^{(j)}; \mathbf{V}^{(j)}), \tag{3.57}$$

where (a) is because of the Markov chain $\widetilde{\mathbf{Z}}^{(j)} \to \mathbf{V}^{(j)} \to \widetilde{\mathbf{Z}}^{(j+1:B)}$. By steps similar to those in (3.21)-(3.24), we can show that

$$I\left(\widetilde{\mathbf{Z}}^{(j)};\mathbf{V}^{(j)}\right)\leqslant\eta_2\delta_2. \tag{3.58}$$

Therefore,

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}^{(1:B)}}\|Q_0^{\otimes Bm\ell}\big)\leqslant\delta+\mathcal{O}(\frac{1}{m}). \tag{3.59}$$

We now show that the covert message and key throughputs are close to the covert capacity. The number of transmitted bits is given by

$$B\left(\log M_U+\log M_V\right)\geqslant B\ell\left(\mathbb{D}(P_1\|P_0)-\frac{1}{m}\chi_2(P_1\|P_0)-\varepsilon\right). \tag{3.60}$$

Hence, the covert rate is given by

$$\begin{aligned}\frac{B\left(\log M_U+\log M_V\right)}{\sqrt{B\ell m\delta}}&\geqslant\sqrt{\frac{B\ell}{m\delta}}\left(\mathbb{D}(P_1\|P_0)-\frac{1}{m}\chi_2(P_1\|P_0)-\varepsilon\right) &(3.61)\\ &=\sqrt{\frac{2}{\chi_2(Q_1\|Q_0)}}\left(\mathbb{D}(P_1\|P_0)-\frac{1}{m}\chi_2(P_1\|P_0)-\varepsilon\right). &(3.62)\end{aligned}$$

The number of key bits used is given by

$$\begin{aligned}&\log M_K+(B-1)(\log M_K-\log M_V)^+\\ &\quad=\ell\sum_{i=1}^{q}R_{K,i}+(B-1)\ell\left(\sum_{i=1}^{q}(R_{K,i}-R_{V,i})\right)^+ &(3.63)\\ &\quad\leqslant\ell\sum_{i=1}^{q}I(X_i;\widetilde{Z}|X_{i+1:q})\\ &\qquad+(B-1)\ell\left(\sum_{i=1}^{q}\Big(I(X_i;\widetilde{Z}|X_{i+1:q})-I(X_i;\widetilde{Y}|X_{i+1:q})+2\varepsilon/q\Big)\right)^+ &(3.64)\\ &\quad=\ell I(\widetilde{X};\widetilde{Z})+(B-1)\ell\left(I(\widetilde{X};\widetilde{Z})-I(\widetilde{X};\widetilde{Y})+2\varepsilon\right)^+ &(3.65)\end{aligned}$$

$$\leqslant \ell \mathbb{D}(Q_1 \| Q_0) + (B-1)\ell \left(\mathbb{D}(Q_1 \| Q_0) - \mathbb{D}(P_1 \| P_0) + \frac{1}{m} \chi_2(P_1 \| P_0) + 2\varepsilon \right)^+ \tag{3.66}$$

Therefore, the key throughput is

$$\frac{\log M_K + (B-1)(\log M_K - \log M_V)^+}{\sqrt{B\ell m\delta}} \leqslant \sqrt{\frac{2}{\chi_2(Q_1 \| Q_0)}} \left(\mathbb{D}(Q_1 \| Q_0) - \mathbb{D}(P_1 \| P_0) + \frac{1}{m} \chi_2(P_1 \| P_0) + 2\varepsilon \right)^+ + \sqrt{\frac{2}{\chi_2(Q_1 \| Q_0)}} \frac{\mathbb{D}(Q_1 \| Q_0)}{B}, \tag{3.67}$$

and the last term vanishes for large B. This shows that the covert message and key throughputs are close to the covert capacity.

3.3.6 Degraded case

We now show that when Willie's channel is degraded w.r.t. Bob's channel, we do not require any chaining or any key. We show that using following proposition.

Proposition 5. *The MLC rates*

$$R_{U,i} = I(X_i; \widetilde{Y} | X_{i+1:q}) - \frac{\varepsilon}{q}, \quad R_{V,i} = 0, \quad R_{K,i} = 0 \tag{3.68}$$

satisfy both reliability and resolvability constraints when Willie's channel is degraded w.r.t. Bob's channel.

Proof. Using the same arguments as for the general case, we can show that for any $\mathcal{S} \subseteq [\![1, q]\!]$,

$$\sum_{i \in \mathcal{S}} (R_{U,i} + R_{V,i}) \leqslant I(X_{\mathcal{S}}, \widetilde{Y} | X_{\mathcal{S}^c}) - \frac{|\mathcal{S}|\varepsilon}{q}. \tag{3.69}$$

Hence, the rates satisfy the constraints for reliability given in (3.7).

When Willie's channel is degraded w.r.t. Bob's channel, we have

$$W_{Z|X}(z|x) = \sum_{y} W_{Y|X}(y|x) W_{Z|Y}(z|y). \tag{3.70}$$

The transition probability of the super channel is given by

$$\begin{aligned}
\widetilde{W}_{\widetilde{Z}|\widetilde{X}}(\widetilde{z}|\widetilde{x}) &= \prod_{j=1}^{m} W(z_j|x_j) && (3.71)\\
&= \prod_{j=1}^{m} \sum_{y_j} W(y_j|x_j) W(z_j|y_j) && (3.72)\\
&= \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \prod_{j=1}^{m} W(y_j|x_j) W(z_j|y_j) && (3.73)\\
&= \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \widetilde{W}(\widetilde{y}|\widetilde{x}) \widetilde{W}(\widetilde{z}|\widetilde{y}). && (3.74)
\end{aligned}$$

Hence, the super channel corresponding to Willie's channel is degraded w.r.t. the super channel corresponding to Bob's channel.

The channel corresponding to the i-th level of MLC for Willie's channel is given by:

$$\begin{aligned}
W_{\widetilde{Z},X_{i+1:q}|X_i}(\widetilde{z}, x_{i+1:q}|x_i) &= \frac{\sum_{x_{1:i-1}} P_X^{\otimes q}(x_{1:q}) \widetilde{W}_{\widetilde{Z}|\widetilde{X}}(\widetilde{z}|\widetilde{x}(x_{1:q}))}{P_X(x_i)} && (3.75)\\
&= \frac{\sum_{x_{1:i-1}} P_X^{\otimes q}(x_{1:q}) \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \widetilde{W}(\widetilde{y}|\widetilde{x}(x_{1:q})) \widetilde{W}(\widetilde{z}|\widetilde{y})}{P_X(x_i)} && (3.76)\\
&= \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \widetilde{W}(\widetilde{z}|\widetilde{y}) \frac{\sum_{x_{1:i-1}} P_X^{\otimes q}(x_{1:q}) \widetilde{W}(\widetilde{y}|\widetilde{x}(x_{1:q}))}{P_X(x_i)} && (3.77)\\
&= \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \widetilde{W}(\widetilde{z}|\widetilde{y}) W_{\widetilde{Y},X_{i+1:q}|X_i}(\widetilde{y}, x_{i+1:q}|x_i). && (3.78)
\end{aligned}$$

This shows that the channels corresponding to each levels of MLC for Willie's channel are

degraded w.r.t. those of Bob's channel. Hence, we have

$$I(X_i;\widetilde{Y}|X_{i+1:q}) = I(X_i;\widetilde{Y},X_{i+1:q}) > I(X_i;\widetilde{Z},X_{i+1:q}) = I(X_i;\widetilde{Z}|X_{i+1:q}). \tag{3.79}$$

We choose ε such that $I(X_i;\widetilde{Y},X_{i+1:q}) - \varepsilon/q > I(X_i;\widetilde{Z},X_{i+1:q}) + \varepsilon/q$. Then, for every $\mathcal{S} \subseteq [\![1,q]\!]$, we can bound the sum-rates as follows:

$$\begin{aligned}
\sum_{i\in\mathcal{S}}(R_{U,i}+R_{K,i}) &= \sum_{i\in\mathcal{S}}\left(I(X_i;\widetilde{Y}|X_{i+1:q}) - \frac{\varepsilon}{q}\right) && (3.80)\\
&\geqslant \sum_{i\in\mathcal{S}}\left(I(X_i;\widetilde{Z}|X_{i+1:q}) + \frac{\varepsilon}{q}\right) && (3.81)\\
&= \sum_{i\in\mathcal{S}} I(X_i;\widetilde{Z}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}},\{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}) + \frac{|\mathcal{S}|\varepsilon}{q} && (3.82)\\
&= \sum_{i\in\mathcal{S}}\Big(I(X_i;\widetilde{Z},\{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}})\\
&\qquad - I(X_i;\{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}})\Big) + \frac{|\mathcal{S}|\varepsilon}{q} && (3.83)\\
&\overset{(a)}{=} \sum_{i\in\mathcal{S}} I(X_i;\widetilde{Z},\{X_j\}_{j\in[\![i+1,q]\!]\setminus\mathcal{S}}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) + \frac{|\mathcal{S}|\varepsilon}{q} && (3.84)\\
&\geqslant \sum_{i\in\mathcal{S}} I(X_i;\widetilde{Z}|\{X_j\}_{j\in[\![i+1,q]\!]\cap\mathcal{S}}) + \frac{|\mathcal{S}|\varepsilon}{q} && (3.85)\\
&= I(X_{\mathcal{S}};\widetilde{Z}) + \frac{|\mathcal{S}|\varepsilon}{q}, && (3.86)
\end{aligned}$$

where (a) follows from the independence of $\{X_i\}$. This shows that the MLC rates in (3.68) satisfy the resolvability rate constraints given in (3.11). □

Since we are not using any key, we do not need chaining to achieve optimal rates. Fig.3.2 shows an illustration of the rate region of MLC with two levels for a degraded case.

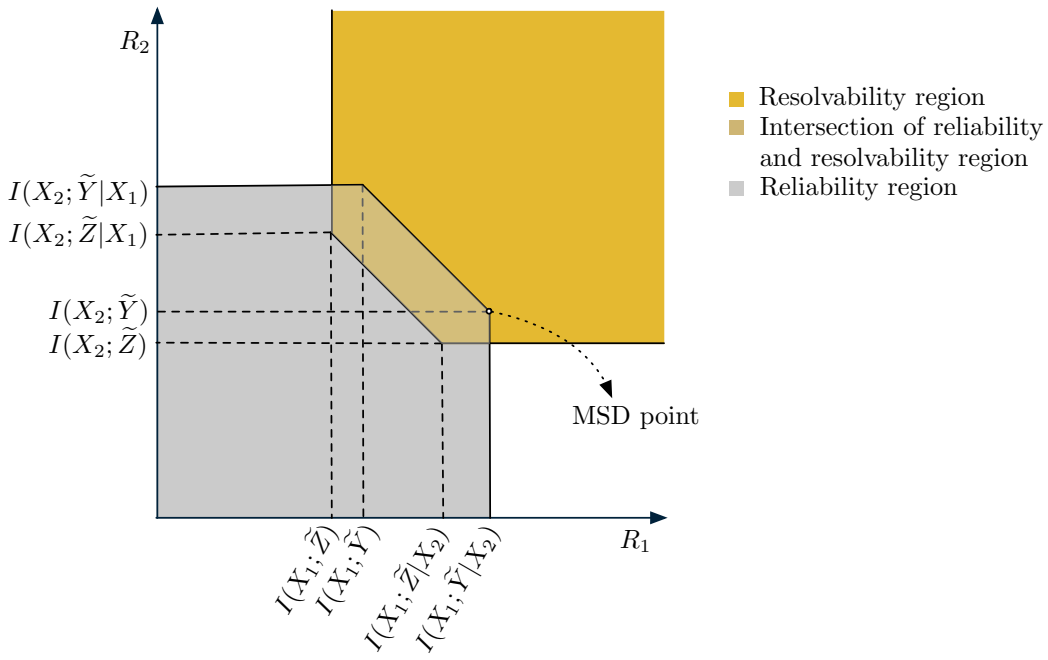

Figure 3.2: Rate region for MLC with two levels with MSD operating point

APPENDIX

3.A Proof of Lemma 3

The probability of error averaged over the codebook distribution is given by

$$\mathbb{E}\left[\mathbb{P}\Big(\widehat{\mathbf{W}} \neq \mathbf{W} | \mathbf{K}=\mathbf{k}\Big)\right] \tag{3.87}$$

$$= \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}}\sum_{\mathbf{w}} \frac{1}{M} \widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k})) \right.$$

$$\left. \mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}}) \notin \mathcal{T}^{\ell}_{\Gamma} \text{ or } \exists \mathbf{w}' \neq \mathbf{w} \text{ s.t. } (\mathbf{X}_{1:q}(\mathbf{w}',\mathbf{k}),\widetilde{\mathbf{y}}) \in \mathcal{T}^{\ell}_{\Gamma}\right\}\right] \tag{3.88}$$

$$= \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}} \widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k})) \right.$$

$$\left. \mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}),\widetilde{\mathbf{y}}) \notin \mathcal{T}^{\ell}_{\Gamma} \text{ or } \exists \mathbf{w} \neq \mathbf{1} \text{ s.t. } (\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}}) \in \mathcal{T}^{\ell}_{\Gamma}\right\}\right] \tag{3.89}$$

$$\leqslant \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}} \widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}))\, \mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}),\widetilde{\mathbf{y}}) \notin \mathcal{T}^{\ell}_{\Gamma}\right\}\right] +$$

$$\sum_{\mathbf{w}\neq\mathbf{1}} \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}} \widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}))\, \mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}}) \in \mathcal{T}^{\ell}_{\Gamma}\right\}\right] \tag{3.90}$$

$$\leqslant \sum_{\mathcal{S}\subseteq[\![1,q]\!]} \mathbb{P}_{\widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}} \prod_{i=1}^{q} P^{\otimes \ell}_{X_i}}\left(\log \frac{\widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}^{\ell}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k})\right)}{P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}\left(\widetilde{Y}^{\ell}|\mathbf{X}_{\mathcal{S}^c}(\mathbf{1},\mathbf{k})\right)} \leqslant \gamma_{\mathcal{S}}\right) +$$

$$\sum_{\mathbf{w}\neq\mathbf{1}} \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}} \widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}))\, \mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}}) \in \mathcal{T}^{\ell}_{\Gamma}\right\}\right]. \tag{3.91}$$

We first analyze the first term on the right hand side of (3.91). We have

$$\log \frac{\widetilde{W}^{\otimes \ell}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{\mathbf{Y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k})\right)}{P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}\left(\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}(\mathbf{1},\mathbf{k})\right)} = \sum_{i=1}^{\ell} \log \frac{\widetilde{W}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}_i|X_{1:q,i}(\mathbf{1},\mathbf{k})\right)}{P_{\widetilde{Y}|X_{\mathcal{S}^c}}\left(\widetilde{Y}_i|X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right)}. \tag{3.92}$$

Let

$$A_i = \log \frac{\widetilde{W}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}_i|X_{1:q,i}(\mathbf{1},\mathbf{k})\right)}{P_{\widetilde{Y}|X_{\mathcal{S}^c}}\left(\widetilde{Y}_i|X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right)}. \tag{3.93}$$

We have

$$\mathbb{E}\left[A_i\right] = I(X_{\mathcal{S}};\widetilde{Y}|X_{\mathcal{S}^c}). \tag{3.94}$$

We now show that A_i is bounded. For any $\mathcal{S} \subseteq [\![1,q]\!]$,

$$P_{\widetilde{Y}|X_{\mathcal{S}^c}}\left(\widetilde{Y}_i|X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right) = \frac{1}{2^{|\mathcal{S}|}} \sum_{x_{\mathcal{S}}} \widetilde{W}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}_i|x_{\mathcal{S}}, X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right) \tag{3.95}$$

$$\geqslant \frac{1}{2^{|\mathcal{S}|}} \widetilde{W}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}_i|X_{\mathcal{S},i}(\mathbf{1},\mathbf{k}), X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right). \tag{3.96}$$

Therefore,

$$A_i \leqslant |\mathcal{S}| \log 2 \leqslant q \log 2. \tag{3.97}$$

Also,

$$\frac{\widetilde{W}_{\widetilde{Y}|X_{1:q}}\left(\widetilde{Y}_i|X_{1:q,i}(\mathbf{1},\mathbf{k})\right)}{P_{\widetilde{Y}|X_{\mathcal{S}^c}}\left(\widetilde{Y}_i|X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k})\right)} = \frac{P_1(Y_{i,\mathcal{A}^q(X_{1:q,i}(\mathbf{1},\mathbf{k}))}) \prod_{h' \neq \mathcal{A}^q(X_{1:q,i}(\mathbf{1},\mathbf{k}))} P_0(Y_{i,h'})}{\frac{1}{2^{|\mathcal{S}|}} \sum_{j \in \mathcal{A}^q(X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k}))} P_1(Y_{i,j}) \prod_{h \neq j} P_0(Y_{i,h})} \tag{3.98}$$

$$= \frac{\frac{P_1(Y_{i,j'})}{P_0(Y_{i,j'})}}{\frac{1}{2^{|\mathcal{S}|}} \sum_{j \in \mathcal{A}^q(X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k}))} \frac{P_1(Y_{i,j})}{P_0(Y_{i,j})}} \tag{3.99}$$

$$\geqslant \frac{P_1(Y_{i,j'})}{\frac{1}{2^{|\mathcal{S}|}}\sum_{j\in\mathcal{A}^q(X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k}))}\frac{1}{P_0(Y_{i,j})}} \tag{3.100}$$

$$\geqslant \frac{\mu_1}{\frac{1}{2^{|\mathcal{S}|}}\sum_{j\in\mathcal{A}^q(X_{\mathcal{S}^c,i}(\mathbf{1},\mathbf{k}))}\frac{1}{\mu_0}} \tag{3.101}$$

$$\geqslant \mu_0\mu_1, \tag{3.102}$$

where

$$\mu_0 = \min_{y\in\text{supp}(P_0)} P_0(y), \qquad \mu_1 = \min_{y\in\text{supp}(P_1)} P_1(y).$$

From (3.97) and (3.102), we have

$$\log(\mu_0\mu_1) \leqslant A_i \leqslant q\log 2 \implies |A_i| \leqslant \max\left(\log\frac{1}{\mu_0\mu_1}, q\log 2\right). \tag{3.103}$$

By choosing $\gamma_{\mathcal{S}} = \ell\left(I(X_{\mathcal{S}};\widetilde{Y}|X_{\mathcal{S}^c}) - \epsilon/q\right)$ and using Hoeffding's inequality, for large enough q, we have

$$\mathbb{P}\left(\log\frac{\widetilde{W}^{\otimes\ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{x}_1,\ldots,\mathbf{x}_q)}{P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{S}^c})} \leqslant \gamma_{\mathcal{S}}\right) \leqslant e^{-\frac{\ell\epsilon^2}{2(\log 2)^2q^4}}. \tag{3.104}$$

We now analyze the second term in (3.91). For any $\mathbf{w}\neq\mathbf{1}$, we let $\mathcal{G}_{\mathbf{w}}\subset[\![1,q]\!]$ denote the set of levels j for which $\mathbf{w}_j = (u_j,v_j) = \mathbf{1}$. Then, we have

$$\begin{aligned}
&\mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}}\widetilde{W}^{\otimes\ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}))\,\mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}})\in\mathcal{T}^{\ell}_{\Gamma}\right\}\right] \\
&\quad= \sum_{\widetilde{\mathbf{y}}}\sum_{\{\mathbf{x}_j(\mathbf{1},k_j)\}_{j\in\mathcal{G}_{\mathbf{w}}}}\prod_{j\in\mathcal{G}_{\mathbf{w}}}P^{\otimes\ell}_{X_j}(\mathbf{x}_j(\mathbf{1},k_j))\sum_{\{\mathbf{x}_{j'}(\mathbf{1},k_{j'}),\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'})\}_{j'\in\mathcal{G}^c_{\mathbf{w}}}} \\
&\qquad\prod_{j'\in\mathcal{G}^c_{\mathbf{w}}}P^{\otimes\ell}_{X_{j'}}(\mathbf{x}_{j'}(\mathbf{1},k_{j'}))P^{\otimes\ell}_{X_{j'}}(\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'}))\;\widetilde{W}^{\otimes\ell}_{\widetilde{Y}|X_{1:q}}(\widetilde{\mathbf{y}}|\mathbf{x}_{1:q}(\mathbf{1},\mathbf{k})) \\
&\qquad\qquad\mathbb{1}\left\{(\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}),\mathbf{x}_{\mathcal{G}^c_{\mathbf{w}}}(\mathbf{w},\mathbf{k}))\in\mathcal{T}^{\ell}_{\Gamma}\right\}
\end{aligned} \tag{3.105}$$

$$= \sum_{\widetilde{\mathbf{y}}} \sum_{\{\mathbf{x}_j(\mathbf{1},k_j)\}_{j\in\mathcal{G}_{\mathbf{w}}}} \prod_{j\in\mathcal{G}_{\mathbf{w}}} P_{X_j}^{\otimes\ell}(\mathbf{x}_j(\mathbf{1},k_j)) \sum_{\{\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'})\}_{j'\in\mathcal{G}_{\mathbf{w}}^c}} \prod_{j'\in\mathcal{G}_{\mathbf{w}}^c} P_{X_{j'}}^{\otimes\ell}(\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'}))$$
$$P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{G}_{\mathbf{w}}}}(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}))\,\mathbb{1}\left\{(\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}),\mathbf{x}_{\mathcal{G}_{\mathbf{w}}^c}(\mathbf{w},\mathbf{k}))\in\mathcal{T}_\Gamma^\ell\right\} \tag{3.106}$$

$$\overset{(a)}{\leqslant} e^{-\gamma_{\mathcal{G}_{\mathbf{w}}^c}} \sum_{\widetilde{\mathbf{y}}} \sum_{\{\mathbf{x}_j(\mathbf{1},k_j)\}_{j\in\mathcal{G}_{\mathbf{w}}}} \prod_{j\in\mathcal{G}_{\mathbf{w}}} P_{X_j}^{\otimes\ell}(\mathbf{x}_j(\mathbf{1},k_j)) \sum_{\{\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'})\}_{j'\in\mathcal{G}_{\mathbf{w}}^c}}$$
$$\prod_{j'\in\mathcal{G}_{\mathbf{w}}^c} P_{X_{j'}}^{\otimes\ell}(\mathbf{x}_{j'}(\mathbf{w}_{j'},k_{j'}))\widetilde{W}_{\widetilde{Y}|X_{1:q}}^{\otimes\ell}\left(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}),\mathbf{x}_{\mathcal{G}_{\mathbf{w}}^c}(\mathbf{w}_{\mathcal{G}_{\mathbf{w}}^c},\mathbf{k})\right) \tag{3.107}$$

$$= e^{-\gamma_{\mathcal{G}_{\mathbf{w}}^c}}, \tag{3.108}$$

where (a) follows since $P_{\widetilde{\mathbf{Y}}|\mathbf{X}_{\mathcal{G}_{\mathbf{w}}}}(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k})) \leqslant e^{-\gamma_{\mathcal{G}_{\mathbf{w}}^c}}\widetilde{W}_{\widetilde{Y}|X_{1:q}}^{\otimes\ell}\left(\widetilde{\mathbf{y}}|\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}),\mathbf{x}_{\mathcal{G}_{\mathbf{w}}^c}(\mathbf{w},\mathbf{k})\right)$ for $(\mathbf{x}_{\mathcal{G}_{\mathbf{w}}}(\mathbf{1},\mathbf{k}),\mathbf{x}_{\mathcal{G}_{\mathbf{w}}^c}(\mathbf{w},\mathbf{k}))\in\mathcal{T}_\Gamma^\ell$ and upper bounding the indicator function by 1.

The sum over $\mathbf{w}\neq\mathbf{1}$ in (3.91) can be expressed as the sum over $\mathcal{G}\subset[\![1,q]\!]$ by counting the number of $\mathbf{w}\neq\mathbf{1}$ for which $\mathbf{w}_j=\mathbf{1}$ for $j\in\mathcal{G}$ and $\mathbf{w}_j\neq\mathbf{1}$ for $j\in\mathcal{G}^c$ as follows.

$$\sum_{\mathbf{w}\neq\mathbf{1}} \mathbb{E}\left[\sum_{\widetilde{\mathbf{y}}} \widetilde{W}_{\widetilde{Y}|X_{1:q}}^{\otimes\ell}(\widetilde{\mathbf{y}}|\mathbf{X}_{1:q}(\mathbf{1},\mathbf{k}))\,\mathbb{1}\left\{(\mathbf{X}_{1:q}(\mathbf{w},\mathbf{k}),\widetilde{\mathbf{y}})\in\mathcal{T}_\Gamma^\ell\right\}\right] \tag{3.109}$$
$$\leqslant \sum_{\mathbf{w}\neq\mathbf{1}} e^{-\gamma_{\mathcal{G}_{\mathbf{w}}^c}} \tag{3.110}$$
$$= \sum_{\mathcal{G}\subset[\![1,q]\!]} \prod_{i\in\mathcal{G}^c} (M_{U,i}M_{V,i}-1)e^{-\gamma_{\mathcal{G}^c}} \tag{3.111}$$
$$\leqslant \sum_{\mathcal{G}\subset[\![1,q]\!]} e^{-\gamma_{\mathcal{G}^c}} \prod_{i\in\mathcal{G}^c} M_{U,i}M_{V,i}. \tag{3.112}$$

Therefore, from (3.91), (3.104), and (3.112), we get

$$\mathbb{E}\left[\mathbb{P}\Big(\widehat{\mathbf{W}}\neq\mathbf{W}\Big)\right] \leqslant \sum_{\mathcal{S}\subseteq[\![1,q]\!]} e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}} + \sum_{\mathcal{G}\subset[\![1,q]\!]} e^{-\gamma_{\mathcal{G}^c}} \prod_{i\in\mathcal{G}^c} M_{U,i}M_{V,i} \tag{3.113}$$

$$= \sum_{\mathcal{S}\subseteq[\![1,q]\!]} e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}} + \sum_{\mathcal{G}\subset[\![1,q]\!]} e^{-\ell\left(I(X_{\mathcal{G}^c};\widetilde{Y}|X_{\mathcal{G}})-\frac{\epsilon}{q}-\sum_{i\in\mathcal{G}^c}(R_{U,i}+R_{V,i})\right)} \tag{3.114}$$

$$\overset{(a)}{\leqslant} \sum_{\mathcal{S}\subseteq[\![1,q]\!]} e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}} + \sum_{\mathcal{G}\subset[\![1,q]\!]} e^{-\frac{\ell(\varepsilon-\epsilon)}{q}} \tag{3.115}$$

$$\overset{(b)}{<} m\left(e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}} + e^{-\frac{\ell\epsilon}{q}}\right), \tag{3.116}$$

where

(a) follows from choosing $\sum_{i\in\mathcal{G}^c}(R_{U,i}+R_{V,i}) \leqslant I(X_{\mathcal{G}^c};\widetilde{Y}|X_{\mathcal{G}}) - \varepsilon/q$.

(b) follows by upper bounding the number of subsets of $[\![1,q]\!]$ and by choosing $\varepsilon > 2\epsilon$.

For $\mathcal{G} \subset [\![1,q]\!]$, let $i \in \mathcal{G}^c$. Then, we have

$$I(X_{\mathcal{G}^c};\widetilde{Y}|X_{\mathcal{G}}) = I(X_{\mathcal{G}^c\setminus\{i\}};\widetilde{Y}|X_{\mathcal{G}}) + I(X_i;\widetilde{Y}|X_{[\![1,q]\!]\setminus\{i\}}) \geqslant I(X_i;\widetilde{Y}|X_{[\![1,q]\!]\setminus\{i\}}). \tag{3.117}$$

Because X_i's are independent and uniformly distributed, the quantity $I(X_i;\widetilde{Y}|X_{[\![1,q]\!]\setminus\{i\}})$ is same for all i. Note that given the realizations of $X_{[\![1,q]\!]\setminus\{i\}} = x_{[\![1,q]\!]\setminus\{i\}}$, the position of symbol "1" in the PPM symbol $\widetilde{x}(X_i, x_{[\![1,q]\!]\setminus\{i\}})$ is given by one of the two positions indexed by $\mathcal{A}^q(x_{[\![1,q]\!]\setminus\{i\}})$. Therefore, $I(X_i;\widetilde{Y}|X_{[\![1,q]\!]\setminus\{i\}})$ is equivalent to the mutual information between the input and outputs when using PPM symbols of order 2 in the MLC-PPM scheme, and using [44, Eq.(13)], we obtain

$$I(X_i;\widetilde{Y}|X_{[\![1,q]\!]\setminus\{i\}}) = \mathbb{D}\left(P_1\|P_0\right) - \mathbb{D}\left(P_{\text{PPM}}^2\|P_0^{\otimes 2}\right), \tag{3.118}$$

where P_{PPM}^2 represents the output distribution of Bob's channel when the input is uniform over PPM symbols of order 2. Hence, we can find rates satisfying $\sum_{i\in\mathcal{G}^c}(R_{U,i}+R_{V,i}) \leqslant I(X_{\mathcal{G}^c};\widetilde{Y}|X_{\mathcal{G}}) - \varepsilon/q$ for $\varepsilon/q < \mathbb{D}\left(P_1\|P_0\right) - \mathbb{D}\left(P_{\text{PPM}}^2\|P_0^{\otimes 2}\right)$ for all $\mathcal{G} \subset [\![1,q]\!]$.

3.B Proof of Lemma 4

We denote Q_{PPM}^m by $Q_{\widetilde{Z}}$ and $(Q_{\text{PPM}}^m)^{\otimes \ell}$ by $Q_{\widetilde{\mathbf{Z}}}$. We have

$$Q_{\widetilde{Z}}(\widetilde{z}) = \frac{1}{2^q} Q_0^{\otimes 2^q}(\widetilde{z}) \sum_{j=1}^{2^q} \frac{Q_1(z_j)}{Q_0(z_j)}, \tag{3.119}$$

$$Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}) = \frac{1}{2^{q\ell}} \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \sum_{j=1}^{2^q} \frac{Q_1(z_{i,j})}{Q_0(z_{i,j})}, \tag{3.120}$$

$$Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}}}(\widetilde{\mathbf{z}}|\mathbf{x}_{\mathcal{S}}) = \frac{1}{2^{|\mathcal{S}^c|\ell}} \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \sum_{j \in \mathcal{A}(x_{\mathcal{S},i})} \frac{Q_1(z_{i,j})}{Q_0(z_{i,j})}. \tag{3.121}$$

We define a typical set as follows:

$$\mathcal{T}_\epsilon^\ell(X_{1:q}, \widetilde{Z}) \triangleq \Bigg\{ (\mathbf{x}_1, \ldots, \mathbf{x}_q, \widetilde{\mathbf{z}}) \in \mathcal{X}^\ell \times \cdots \times \mathcal{X}^\ell \times \widetilde{\mathcal{Z}}^\ell : \frac{1}{\ell} \log \frac{Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}}}(\widetilde{\mathbf{z}}|\mathbf{x}_{\mathcal{S}})}{Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} < I(X_{\mathcal{S}}; \widetilde{Z}) + \frac{\epsilon}{q}, \ \forall \mathcal{S} \subseteq [\![1, q]\!] \Bigg\}. \tag{3.122}$$

We denote the expectation over all the random variables except $\mathbf{X}_{1:q}(\mathbf{u}, \mathbf{v}, \mathbf{k})$ by $\mathbb{E}_{\sim \mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})}$. Then, the expectation of relative entropy over the codebook distribution is

$$\mathbb{E}\left[\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}|\mathbf{V}=\mathbf{v}} \| Q_{\widetilde{\mathbf{Z}}}\Big)\right] = \mathbb{E}\left[\sum_{\widetilde{\mathbf{z}}} \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}, \mathbf{v}, \mathbf{k})) \log\left(\frac{\sum_{\mathbf{u}',\mathbf{k}'} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}', \mathbf{v}, \mathbf{k}'))}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right)\right] \tag{3.123}$$

$$= \sum_{\widetilde{\mathbf{z}}} \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \mathbb{E}_{\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})}\left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}, \mathbf{v}, \mathbf{k})) \mathbb{E}_{\sim \mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})}\left[\log\left(\frac{\sum_{\mathbf{u}',\mathbf{k}'} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}', \mathbf{v}, \mathbf{k}'))}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right)\right]\right] \tag{3.124}$$

$$\overset{(a)}{\leqslant} \sum_{\widetilde{\mathbf{z}}} \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \mathbb{E}_{\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})} \left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})) \right.$$
$$\left. \log \left(\frac{\sum_{\mathbf{u}',\mathbf{k}'} \mathbb{E}_{\sim \mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})} \left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}',\mathbf{v},\mathbf{k}')) \right]}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} \right) \right] \tag{3.125}$$

$$\overset{(b)}{=} \sum_{\widetilde{\mathbf{z}}} \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \mathbb{E}_{\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})} \left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})) \right.$$
$$\left. \log \left(\frac{\sum_{\mathcal{S} \subseteq [\![1,q]\!]} \prod_{i \in \mathcal{S}} (M_{U,i} M_{K,i} - 1) Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{z}}|\mathbf{X}_{\mathcal{S}^c}(\mathbf{u},\mathbf{v},\mathbf{k}))}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} \right) \right] \tag{3.126}$$

$$\leqslant \sum_{\widetilde{\mathbf{z}}} \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \mathbb{E}_{\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})} \left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})) \right.$$
$$\left. \log \left(1 + \frac{\sum_{\mathcal{S} \subset [\![1,q]\!]} \prod_{i \in \mathcal{S}} (M_{U,i} M_{K,i} - 1) Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{z}}|\mathbf{X}_{\mathcal{S}^c}(\mathbf{u},\mathbf{v},\mathbf{k}))}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} \right) \right] \tag{3.127}$$

$$= \frac{1}{M_U M_K} \sum_{\mathbf{u},\mathbf{k}} \sum_{\widetilde{\mathbf{z}}} \sum_{\mathbf{x}_{1:q}} \frac{1}{2^{q\ell}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q})$$
$$\log \left(1 + \frac{\sum_{\mathcal{S} \subset [\![1,q]\!]} \prod_{i \in \mathcal{S}} (M_{U,i} M_{K,i} - 1) Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{z}}|\mathbf{x}_{\mathcal{S}^c})}{M_U M_K Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} \right) \tag{3.128}$$

$$\overset{(c)}{\leqslant} \sum_{(\mathbf{x}_{1:q},\widetilde{\mathbf{z}}) \in \mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})} \frac{1}{2^{q\ell}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q})$$
$$\log \left(1 + \sum_{\mathcal{S} \subset [\![1,q]\!]} e^{-\ell(\sum_{i \in \mathcal{S}^c}(R_{U,i}+R_{K,i}) - I(X_{\mathcal{S}^c};\widetilde{Z}) - \epsilon/q)} \right)$$
$$+ \sum_{(\mathbf{x}_{1:q},\widetilde{\mathbf{z}}) \notin \mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})} \frac{1}{2^{q\ell}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}) \log \left(1 + 2^q \mu_z^{-m\ell} \right) \tag{3.129}$$

$$\overset{(d)}{\leqslant} \sum_{(\mathbf{x}_{1:q},\widetilde{\mathbf{z}}) \in \mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})} \frac{1}{2^{q\ell}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}) \sum_{\mathcal{S} \subset [\![1,q]\!]} e^{-\ell(\sum_{i \in \mathcal{S}^c}(R_{U,i}+R_{K,i}) - I(X_{\mathcal{S}^c};\widetilde{Z}) - \epsilon/q)}$$
$$+ \sum_{(\mathbf{x}_{1:q},\widetilde{\mathbf{z}}) \notin \mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})} \frac{1}{2^{q\ell}} \widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}) \frac{m\ell 2^{q/m\ell}}{\mu_z} \tag{3.130}$$

$$\leqslant \sum_{\mathcal{S}\subset[\![1,q]\!]} e^{-\ell(\sum_{i\in\mathcal{S}}(R_{U,i}+R_{K,i})-I(X_{\mathcal{S}};\widetilde{Z})-\epsilon/q)} + \frac{m\ell 2^{q/m\ell}}{\mu_z}\mathbb{P}\Big((\mathbf{X}_{1:q},\widetilde{\mathbf{Z}})\notin\mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})\Big) \tag{3.131}$$

$$\overset{(e)}{\leqslant} me^{-\ell(\varepsilon-\epsilon)/q} + \frac{m\ell 2^{q/m\ell}}{\mu_z}\mathbb{P}\Big((\mathbf{X}_{1:q},\widetilde{\mathbf{Z}})\notin\mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})\Big), \tag{3.132}$$

where

(a) follows from Jensen's inequality.

(b) follows from expressing the summation over all $(\mathbf{u}',\mathbf{k}')$ as a summation over $\mathcal{S}\subseteq[\![1,q]\!]$ and

$$(\mathbf{u}',\mathbf{k}')\in\left\{(\mathbf{u}'',\mathbf{k}''):(u_j'',k_j'')=(u_j,k_j)\forall j\in\mathcal{S}^c \text{ and } (u_j'',k_j'')\neq(u_j,k_j)\forall j\in\mathcal{S}\right\},$$

and noting that

$$\mathbb{E}_{\sim\mathbf{X}_{1:q}(\mathbf{u},\mathbf{v},\mathbf{k})}\left[\widetilde{W}_{\widetilde{Z}|X_{1:q}}(\widetilde{\mathbf{z}}|\mathbf{X}_{1:q}(\mathbf{u}',\mathbf{v},\mathbf{k}'))\right]=Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{z}}|\mathbf{X}_{\mathcal{S}^c}(\mathbf{u},\mathbf{v},\mathbf{k}))$$

for all $(\mathbf{u}',\mathbf{k}')$ in that set.

(c) follows because for $(\mathbf{x}_{1:q},\widetilde{\mathbf{z}})\in\mathcal{T}_\epsilon^\ell(X_{1:q},\widetilde{Z})$ and $\mathcal{S}\subseteq[\![1,q]\!]$,

$$\frac{Q_{\widetilde{\mathbf{Z}}|\mathbf{X}_{\mathcal{S}^c}}(\widetilde{\mathbf{z}}|\mathbf{x}_{\mathcal{S}^c})}{Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\leqslant e^{\ell(I(X_{\mathcal{S}^c};\widetilde{Z})+\epsilon/q)}.$$

(d) follows because $\log(1+x^m)\leqslant mx$ for $x\geqslant 0$.

(e) follows because $\sum_{i\in\mathcal{S}}(R_{U,i}+R_{K,i})\geqslant I(X_{\mathcal{S}};\widetilde{Z})+\varepsilon/q$.

Using Hoeffding's inequality as in the previous section, we can show that

$$\mathbb{P}\Big((\mathbf{X}_{1:q}, \widetilde{\mathbf{Z}}) \notin \mathcal{T}_\epsilon^\ell(X_{1:q}, \widetilde{Z})\Big) \leqslant m e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}}. \tag{3.133}$$

Therefore, for $0 < \epsilon < \varepsilon/2$, we have

$$\mathbb{E}\left[\mathbb{D}\Big(P_{\widetilde{\mathbf{Z}}|\mathbf{V}=\mathbf{v}} \| Q_{\widetilde{\mathbf{Z}}}\Big)\right] \leqslant m e^{-\ell\epsilon/q} + \frac{m^2 \ell 2^{q/m\ell}}{\mu_z} e^{-\frac{\ell\epsilon^2}{2(\log 2)^2 q^4}}. \tag{3.134}$$

CHAPTER 4
TOWARDS PRACTICAL CODES FOR COVERT COMMUNICATION OVER BI-DMCs

In this chapter, we discuss the design of practical codes for covert communication over a BI-DMC using MLC. The code design, in general, involves using keys and chaining over multiple blocks. To simplify the analysis, we discuss code design for the case in which the warden's channel is degraded w.r.t. Bob's channel. We can generalize the code design for non-degraded cases using a chaining scheme as explained in the previous chapter. First, we show that the equivalent channel corresponding to a particular level when using MSD remains unchanged when we alter the number of levels, which tremendously simplifies the code design. After establishing additional properties of the equivalent channels, we analyze the probability of error at Bob's decoder and covertness at Willie's receiver. Towards the end of this chapter, we will discuss an explicit low-complexity code construction using polar codes and invertible extractors.

4.1 Equivalent channel for each level

We now prove that the channel corresponding to each level of MLC with MSD for uniformly distributed inputs can be represented by an equivalent channel that does not depend on the number of levels used in the scheme. In Section 3.3, we established that we can consider the i-th channel as a channel with input X_i and output $(\widetilde{Y}, X_{i+1:q})$. Because of the structure of PPM, for a known input to the levels $i+1$ to q, $x_{i+1:q}$, the position of the symbol "1" in the PPM symbol can be narrowed down to 2^i positions. Using the notation defined earlier, $\mathcal{A}^q(x_{i+1:q})$ denotes the set of indices of those positions. According to the PPM mapper defined in Section 3.1, the position of the symbol "1" is in the first half of the PPM symbol if $x_q = 0$ and in the second half otherwise. Similarly, if $x_i = 0$, the position

of the symbol "1" is in the first half of the subset of indices determined by the values of $(x_{i+1}, \ldots, x_q)$ and second half otherwise. Table 4.1 shows an illustration of this mapping for $m = 16$.

Table 4.1: Illustration of PPM mapper for $m = 16$

PPM symbol index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
x_2	0		1		0		1		0		1		0		1	
x_3	0				1				0				1			
x_4	0								1							

The set $\mathcal{A}^q(x_{i+1:q})$ also defines the indices for the aforementioned PPM symbols because of the indexing defined earlier. To estimate the i-th bit, we have to consider only the outputs at the positions with indices in $\mathcal{A}^q(x_{i+1:q})$. Assuming that the inputs to levels 1 to $i-1$ are uniformly distributed, we can express the effective channel as

$$W^i(\widetilde{y}_{\mathcal{A}^q(x_{i+1:q})}|x_i) = \frac{1}{2^{i-1}} \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{\widetilde{y}_{[\![1,2^q]\!]\setminus\mathcal{A}^q(x_{i+1:q})}} W^{\otimes 2^q}\left(\widetilde{y}|\widetilde{x}(x_{1:q})\right) \tag{4.1}$$

$$= \frac{1}{2^{i-1}} \sum_{x_1} \cdots \sum_{x_{i-1}} P_0^{\otimes 2^{i-1}}\left(\widetilde{y}_{\mathcal{A}^q(x_{i+1:q})\setminus\mathcal{A}^q(x_{1:q})}\right) P_1(\widetilde{y}_{\mathcal{A}^q(x_{1:q})}). \tag{4.2}$$

Since the channel is memoryless, once the decoder selects the 2^i positions indexed by $\mathcal{A}^q(x_{i+1:q})$, the distribution of the selected output symbols is independent of the input to the levels $i+1$ to q. Hence, we can represent the above channel equivalently by

$$W^i(y_1, \ldots, y_{2^i}|x_i) = \frac{1}{2^{i-1}} \sum_{x_1} \cdots \sum_{x_{i-1}} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_{\mathcal{A}^i(x_{1:i})})}{P_0(y_{\mathcal{A}^i(x_{1:i})})} \tag{4.3}$$

$$= \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_k)}{P_0(y_k)} \tag{4.4}$$

$$= \frac{1}{2^{i-1}} \sum_{j=2^{i-1}x_i+1}^{2^{i-1}(1+x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_j)}{P_0(y_j)}. \tag{4.5}$$

The crucial aspect of this characterization is showing that this channel remains unchanged

irrespective of the number of levels q used. Note that the above channel is equivalent to the i-th channel in an MLC with i levels.

Lemma 5. *When using MSD, for $i \in [\![1, q]\!]$, the bits of the i-th level of MLC-PPM setup are effectively transmitted over a BI-DMC with transition probability*

$$W^i(y_1, \ldots, y_{2^i}|x_i) = \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_k)}{P_0(y_k)}. \tag{4.6}$$

Moreover, this channel remains unchanged irrespective of the number of levels q used.

Proof. Since, X_i is independent of $X^{i+1:q}$ and the common randomness, we may consider the i-th level as a channel with input X_i and output $(\widetilde{Y}, X^{i+1:q}, C^{i+1:q})$. Because of the structure of PPM, for a known input $x^{i+1:q}$ to the levels $i+1$ to q, the position of the symbol 1 in the PPM symbol can be narrowed down to 2^i positions. Let $\mathcal{A}^q(x^{i+1:q})$ denote the set of indices of those positions. This set also define the indices for the PPM symbol because of the indexing defined earlier. To make a decision about the i-th bit, we need only look at the outputs at the positions $\mathcal{A}^q(x^{i+1:q})$. Since the common randomness of lower levels $C^{1:i-1}$ is not used in the decoding, this effective channel is given by

$$W(\widetilde{y}_{\mathcal{A}^q(x^{i+1:q})}|x_i) = \frac{1}{2^{i-1}} \sum_{x_1} \cdots \sum_{x_{i-1}} W^{\otimes 2^i}\left(\widetilde{y}_{\mathcal{A}^q(x^{i+1:q})}|\widetilde{x}(x^{1:q})\right) \tag{4.7}$$

$$= \frac{1}{2^{i-1}} \sum_{x_1} \cdots \sum_{x_{i-1}} P_0^{\otimes 2^i - 1}\left(\widetilde{y}_{\mathcal{A}^q(x^{i+1:q}) \setminus \mathcal{A}^q(x^{1:q})}\right) P_1(\widetilde{y}_{\mathcal{A}^q(x^{1:q})}). \tag{4.8}$$

Since the channel is memoryless, once the decoder select the 2^i positions indexed by $\mathcal{A}^q(x^{i+1:q})$, the distribution is independent of inputs to the higher levels and we can represent the above channel as in (4.6). □

Since the i-th level channel is constant, we know that its capacity is fixed, and we can characterize it more precisely as follows.

Lemma 6. *For $i \in [\![1, q]\!]$, the capacity C_i of the i-th level satisfies*

$$C_i \leqslant \frac{1}{2^i}\chi_3(P_1\|P_0) - \frac{1}{2^{2i-1}}\left[\chi_2(Q_1||Q_0)^2 - \frac{1}{6}\chi_3(Q_1||Q_0)\right] + \frac{1}{3\times 2^{3(i-1)}}\left[\frac{1}{3}\chi_4(Q_1||Q_0) - \chi_2(Q_1||Q_0)^2\right]. \quad (4.9)$$

Proof. The capacity of the i-th level is $I(X_i;\widetilde{Y}|X_{i+1:q})$ with X_j for $j \in [\![1, q]\!]$ distributed uniformly in $\{0, 1\}$. Since X_i is independent of $X_{i+1:q}$, we have

$$I(X_i;\widetilde{Y}|X_{i+1:q}) = I(X_i;\widetilde{Y}, X_{i+1:q}) \quad (4.10)$$

$$= I(X_{1:i};\widetilde{Y}, X_{i+1:q}) - I(X_{1:i-1};\widetilde{Y}, X_{i:q}) \quad (4.11)$$

$$= I(X_{1:i};\widetilde{Y}|X_{i+1:q}) - I(X_{1:i-1};\widetilde{Y}|X_{i:q}). \quad (4.12)$$

Similar to the argument used for deriving the equivalent channel of each level, we can show that the quantity$I(X_{1:i};\widetilde{Y}, X_{i+1:q})$ is equivalent to $I(X_{1:i};\widetilde{Y}_{\mathcal{A}^q(X_{i+1:q})})$, which represents the capacity of a PPM channel of order 2^i. Using [44, Eq.(13)], we obtain

$$I(X_1, \ldots, X_i; Y_1, \ldots, Y_{2^i}) = \mathbb{D}(P_1\|P_0) - \mathbb{D}\left(P_{\text{PPM}}^{2^i}\|P_0^{\otimes 2^i}\right), \quad (4.13)$$

where $P_{\text{PPM}}^{2^i}$ represents the output distribution when the input is uniform over all possible PPM symbols of order 2^i. Hence, we have

$$C_i = I(X_i;\widetilde{Y}|X_{i+1:q}) = \mathbb{D}\left(P_{\text{PPM}}^{2^{i-1}}\|P_0^{\otimes 2^{i-1}}\right) - \mathbb{D}\left(P_{\text{PPM}}^{2^i}\|P_0^{\otimes 2^i}\right) \quad (4.14)$$

Upper bounding the right-hand side using [44, Lemma 1] yields the desired result. □

Note that, for higher levels, the capacity C_i goes to zero exponentially in the index of the level. Since the sum of the capacities converges to $\mathbb{D}(P_1\|P_0)$ by [44, Lemma 2] and by the capacity-achieving property of MLC with MSD [48], very few levels concentrate most

of the capacity. As an example, Table 4.2 shows the capacity per level of first 16 levels for a binary symmetric channel with cross-over probability 0.1. Notice that the first 5 levels concentrate 93.4% of the total capacity.

Table 4.2: Capacity of first 16 levels for a BSC with cross-over probability 0.1

Level, i	1	2	3	4	5	6	7	8
Capacity, C_i	0.7421	0.6387	0.4918	0.3214	0.1749	0.0853	0.0413	0.0203
Level, i	9	10	11	12	13	14	15	16
Capacity, C_i	0.0101	0.0050	0.0025	0.0013	0.0006	0.0003	0.0002	0.0001

We state some of the properties of equivalent channels in the following lemmas.

Lemma 7. *The equivalent channel defined in* (4.6) *is symmetric.*

Proof. The equivalent channel W^i has the following property:

$$W^i(y_1, \ldots, y_{2^i}|x_i) = W^i(y_{2^{i-1}+1}, \ldots, y_{2^i}, y_1, \ldots, y_{2^{i-1}}|x_i \oplus 1). \tag{4.15}$$

Therefore, the equivalent channel is symmetric. □

Definition 1. *Let $\mathcal{P}_i$ represent the set of all permutations of $[\![1, 2^i]\!]$. We define the set Π_i as*

$$\Pi_i \triangleq \left\{\sigma \in \mathcal{P}_i : \forall i \in [\![1, 2^{i-1}]\!], \sigma(i) \in [\![1, 2^{i-1}]\!]\right\}.$$

Let Π_i^ℓ represent the set of permutations of a vector of length ℓ, whose components are vectors of length 2^i, such that each component is permuted by one element of Π_i.

Note that for $\pi \in \Pi_i^\ell$ and $\widetilde{\mathbf{y}} \in \widetilde{\mathcal{Y}}_i^\ell$, $\pi(\widetilde{\mathbf{y}})$ is a permutation of $\widetilde{\mathbf{y}}$ such that the components of each $\widetilde{y}_j \in \mathcal{Y}^{2^i}$ are permuted in such a way that the components in the first half remain in the first half.

Lemma 8. *The equivalent channel W^i is invariant under any permutation $\pi \in \Pi_i$.*

Proof. We have

$$W^i\left(y_{\pi(1)},\ldots,y_{\pi(2^i)}|x_i\right) = \frac{1}{2^{i-1}}\sum_{k\in\mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_{\pi(1)},\ldots,y_{\pi(2^i)})\frac{P_1(y_{\pi(k)})}{P_0(y_{\pi(k)})} \tag{4.16}$$

$$\stackrel{(a)}{=} \frac{1}{2^{i-1}}\sum_{k\in\mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1,\ldots,y_{2^i})\frac{P_1(y_k)}{P_0(y_k)} \tag{4.17}$$

$$= W^i(y_1,\ldots,y_{2^i}|x_i), \tag{4.18}$$

where (a) follows because $\mathcal{A}^i(x_i)$ is either $[\![1, 2^{i-1}]\!]$ or $[\![2^{i-1}+1, 2^i]\!]$ depending on the value of x_i, and from the the definition of Π_i, the summation has the same terms with and without the permutation π. Therefore, W^i is invariant under any permutation $\pi \in \Pi_i$. □

4.2 Analysis of MSD

We now prove that we can achieve reliability for the MLC scheme with MSD if we use *independent* reliability codes for the equivalent channel corresponding to each level. Note that the actual channel for each level is different from the equivalent channel defined in (4.6) because of the use of codes for the lower levels instead of uniformly distributed inputs. Let W^i be the equivalent channel for level i as defined in (4.6). Let ϕ_i and ψ_i represent the encoder and decoder for an (ℓ, m_i)-code for the i-th channel with rate $R_i = \frac{m_i}{\ell}$. We define the rate of the MLC scheme as

$$R \triangleq \sum_{i=1}^{q} R_i. \tag{4.19}$$

Encoding: The encoder partitions the message W with ℓR bits into q messages such that the i-th message W_i contains ℓR_i bits. It then uses ϕ_i to encode W_i into an ℓ-bit sequence $\mathbf{X}_i = (X_{i,1}, \cdots, X_{i,\ell})$. Let $X_{i:q,j} \triangleq (X_{i,j}, X_{i+1,j}, \ldots, X_{q,j})$. The PPM mapper maps $X_{1:q,j}$ to $\widetilde{x}(X_{1:q,j})$ for $j \in [\![1, \ell]\!]$ as defined in (3.3) to form a sequence of ℓ PPM symbols, which is transmitted over the channel.

Decoding: Let $\widetilde{\mathbf{Y}} = (\widetilde{Y}_1, \cdots, \widetilde{Y}_\ell)$ be the received sequence, where $\widetilde{Y}_i = (Y_{i,1}, \ldots, Y_{i,m})$. The decoder starts from the level q and decodes the messages successively from level q to 1. To decode the i-th level, we assume that estimates of $W_{i+1}, \ldots, W_q$ are available at the decoder as $\widehat{W}_{i+1}, \cdots, \widehat{W}_q$. The decoder also has the estimates of the inputs to these decoded levels as $\widehat{\mathbf{X}}_j \triangleq \phi(\widehat{W}_j)$. It uses these estimates to form a sequence

$$\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\hat{\mathbf{X}}_{i+1:q})} \triangleq \left(\widetilde{Y}_{1,\mathcal{A}^q(\hat{X}_{i+1:q,1})}, \ldots, \widetilde{Y}_{\ell,\mathcal{A}^q(\hat{X}_{i+1:q,\ell})}\right), \tag{4.20}$$

where $\widetilde{Y}_{k,\mathcal{A}^q(\hat{X}_{i+1:q,k})} \triangleq (Y_{k,t})_{t\in\mathcal{A}^q(\hat{X}_{i+1:q,k})}$. The decoder estimates the message W_i as $\widehat{W}_i \triangleq \psi_i(\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\hat{\mathbf{X}}_{i+1:q})})$. We define the decoding region of message w_i for the decoder ψ_i as $\mathcal{D}_{w_i} \triangleq \{\widetilde{\mathbf{y}} \in \widetilde{\mathcal{Y}}_i^\ell : \psi_i(\widetilde{\mathbf{y}}) = w_i\}$.

Assumption: For $\widetilde{\mathbf{y}} = (\widetilde{y}_1, \ldots, \widetilde{y}_\ell)$, where $\widetilde{y}_j = (y_{j,1}, \ldots, y_{j,2^i})$, and for any permutation $\pi \in \Pi_i^\ell$, we assume that if $\widetilde{\mathbf{y}} \in \mathcal{D}_{w_i}$, then $\pi(\widetilde{\mathbf{y}}) \in \mathcal{D}_{w_i}$.

This assumption is reasonable because there exist efficient decoders such as the successive cancellation decoder for polar codes that have this property.

Lemma 9. *For the successive cancellation decoder of polar codes,* $\widetilde{\mathbf{y}} \in \mathcal{D}_{w_i} \implies \pi(\widetilde{\mathbf{y}}) \in \mathcal{D}_{w_i}$ *for every* $\pi \in \Pi_i^\ell$.

Proof. Assume ℓ is a power of two. Let $\mathbf{u}_{i,1:\ell} = \mathbf{x}_{i,1:\ell}G_\ell$, where G_ℓ is the polar code transform matrix defined in [46]. In successive decoding of polar codes, the decoder produces the decision by computing

$$\mathbb{P}\Big(U_j = u_j | \widetilde{\mathbf{Y}} = \widetilde{\mathbf{y}}, \mathbf{U}_{1:j-1} = \hat{\mathbf{u}}_{1:j-1}\Big) = \frac{\sum_{\mathbf{u}_{j+1:\ell}} p_{U_{1:\ell}}(\hat{\mathbf{u}}_{1:j}, \mathbf{u}_{j:\ell})(W^i)^{\otimes\ell}(\widetilde{\mathbf{y}}|\mathbf{x}_i)}{p_{U_{1:j-1}}(\mathbf{u}_{1:j-1})W^i(\widetilde{\mathbf{y}}|\mathbf{u}_{1:j-1})}, \tag{4.21}$$

where

$$W^i(\widetilde{\mathbf{y}}|\mathbf{u}_{1:j}) = \sum_{u_{j+1}=0}^{1} \cdots \sum_{u_\ell=0}^{1} (W^i)^{\otimes\ell}(\widetilde{\mathbf{y}}|\mathbf{x}_i)p(u_{j+1}, \ldots, u_q). \tag{4.22}$$

From Lemma 8, we have

$$(W^i)^{\otimes \ell}\left(\pi(\widetilde{\mathbf{y}})|\mathbf{x}_i\right) = (W^i)^{\otimes \ell}\left(\widetilde{\mathbf{y}}|\mathbf{x}_i\right). \tag{4.23}$$

Hence, $\mathbb{P}\Big(U_j = u_j|\widetilde{\mathbf{Y}} = \widetilde{\mathbf{y}}, \mathbf{U}_{1:j-1} = \hat{\mathbf{u}}_{1:j-1}\Big)$ is also invariant under the permutation π and the result follows. □

Lemma 10. *Suppose we have a code (ϕ_i, ψ_i) for the channel*

$$W^i(y_1, \ldots, y_{2^i}|x_i) = \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_k)}{P_0(y_k)}, \tag{4.24}$$

with rate R_i and probability of error ϵ_i, such that ψ_i satisfies the above assumption on the decoding region. We can then design a code for the MLC-PPM scheme with rate $R = \sum_{i=1}^{q} R_i$ and probability of error $\epsilon = \sum_{i=1}^{q} \epsilon_i$.

Proof. See Appendix 4.A. □

4.3 Analysis of covertness

We now turn our attention to the covertness properties of the MLC scheme. Instead of dealing with relative entropy, we work with the variational distance $\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big)$ between the distribution $P_{\widetilde{\mathbf{Z}}}$ induced at the output of the PPM super-channel when coding over ℓ PPM symbols of order m, and the distribution $(Q_{\text{PPM}}^m)^{\otimes \ell}$, which is a product distribution over the ℓ uses of the super channel, and Q_{PPM}^m is the output induced by a uniform input distribution on PPM symbols of order m. For $j \in [\![1, q]\!]$, assume that the codebook at level j consists of M_j codewords $\mathcal{C}_i = \{\mathbf{c}(i_j)\}_{i_j=1}^{M_j}$. Upon denoting the super channel transition probability $\widetilde{W}_{\widetilde{Z}|\widetilde{X}}$ by $\widetilde{W}$, we introduce the distribution

$$P_{\widetilde{\mathbf{Z}}}^{(j)}(\widetilde{\mathbf{z}}) = \frac{1}{2^\ell} \sum_{\mathbf{x_1} \in \{\mathbf{0},\mathbf{1}\}^\ell} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_j} \in \{\mathbf{0},\mathbf{1}\}^\ell} \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1} \in \mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}}|\mathbf{x}_{1:j}, \mathbf{c}_{j+1:q}\right).$$

Intuitively, $P_{\widetilde{\mathbf{Z}}}^{(j)}$ represents the distribution induced at Willie's output when coding all levels from $j+1$ down to q and transmitting uniformly distributed bits on all lower levels. Note that $P_{\widetilde{\mathbf{Z}}}^{(0)} = P_{\widetilde{\mathbf{Z}}}$ and $P_{\widetilde{\mathbf{Z}}}^{(q)} = (Q_{\text{PPM}}^m)^{\otimes \ell}$. Using a triangle inequality repeatedly, we obtain that

$$\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\right) \leqslant \sum_{j=1}^{q} \mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}^{(j-1)}, P_{\widetilde{\mathbf{Z}}}^{(j)}\right). \tag{4.25}$$

Now, we can further upper bound $\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}^{(j-1)}, P_{\widetilde{\mathbf{Z}}}^{(j)}\right)$ as

$$\begin{aligned}
\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}^{(j-1)}, P_{\widetilde{\mathbf{Z}}}^{j}\right) = \sum_{\widetilde{\mathbf{z}}} \Bigg| & \frac{1}{2^\ell} \sum_{\mathbf{x_1}} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_{j-1}}} \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1} \in \mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \\
& \left(\frac{1}{M_j} \sum_{\mathbf{c}_j \in \mathcal{C}_j} \widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}} | \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{c}_j, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_q\right) \right. \\
& \left. - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}} | \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_j, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_q\right) \right) \Bigg| \qquad (4.26) \\
\leqslant \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1} \in \mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}} \Bigg| & \frac{1}{2^\ell} \sum_{\mathbf{x_1}} \cdots \frac{1}{2^\ell} \\
& \sum_{\mathbf{x_{j-1}}} \left(\frac{1}{M_j} \sum_{\mathbf{c}_j \in \mathcal{C}_j} \widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}} | \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{c}_j, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_q\right) \right. \\
& \left. - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}} | \mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_j, \mathbf{c}_{j+1}, \ldots, \mathbf{c}_q\right) \right) \Bigg| \qquad (4.27)
\end{aligned}$$

Notice that the two terms inside the absolute are distributions that only differ in that one has a coded j-th level, while the other has an uncoded j-th level with uniform random bits.

Lemma 11. *For every $j \in [\![1, q]\!]$, consider the channel*

$$W^j(z_1, \ldots, z_{2^j} | x_j) = \frac{1}{2^{j-1}} \sum_{k \in \mathcal{A}^j(x_j)} Q_0^{\otimes 2^j}(z_1, \ldots, z_{2^j}) \frac{Q_1(z_k)}{Q_0(z_k)}. \tag{4.28}$$

Let $P_{\widetilde{\mathbf{Z}}^{(j)}}$ denote the output distribution induced by a code over this channel, and let $Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}$

denote the product output distribution when the input is uniform. If $\mathbb{V}\Big(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}\Big) \leqslant \delta_j$*, then the same code ensures that* $\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}^{(j-1)}, P_{\widetilde{\mathbf{Z}}}^{(j)}\right) \leqslant \delta_j$ *irrespective of the code used for the higher levels. Moreover, for the output of the super channel* $\widetilde{W}_{\widetilde{Z}|X_{1:q}}$*, these codes together ensure* $\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}, (Q_{PPM}^m)^{\otimes \ell}\right) \leqslant \sum_{j=1}^{q} \delta_j$.

Proof. See Appendix 4.B □

Lemma 11 may be viewed as the counterpart of Lemma 10 for resolvability instead of reliability. Since the equivalent channel (4.28) is again invariant with the number of levels $q \geqslant i$, we conclude that we can design channel resolvability codes for a fixed channel while still growing the number of levels with the code length.

To conclude regarding the ability to achieve covertness with the multilevel scheme, we note with calculations similar to [17, (77)-(79)] that

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| Q_0^{\otimes n}\big) \leqslant \mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| (Q_{\text{PPM}}^m)^{\otimes \ell}\big) + 2\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big) \max_{\widetilde{z}} \ell \left| \log \frac{(Q_{\text{PPM}}^m)(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})} \right| + \mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes \ell} \| Q_0^{\otimes n}\big), \quad (4.29)$$

and using [49, (323)]

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| (Q_{\text{PPM}}^m)^{\otimes \ell}\big) \leqslant \ell \log \left(\frac{1}{\min_{\widetilde{z}} (Q_{\text{PPM}}^m)^{\otimes \ell}(\widetilde{z})} \right) \mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big). \quad (4.30)$$

Consequently, provided $\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big)$ decays fast enough at each level and with ℓ scaling as in [44], we may ensure covertness at a throughput close to the covert capacity.

4.4 Towards a concrete polynomial-complexity instantiation

The key observation to instantiate actual codes is that the problem reduces to constructing codes for the equivalent channels identified in Lemma 10 and Lemma 11. If the original channels are degraded, then the equivalent channels are also degraded as shown in Section 3.3.6. Since the successive cancellation decoder of polar codes satisfies the assump-

tion on the decoding regions that we used for the proof of reliability, polar codes would be a suitable solution to achieve the covert capacity. In fact, we know from [39] how to design polar codes simultaneously for reliability and resolvability, with a negligible amount of shared randomness and metrics (probability of error and variational distance) that decay fast with the blocklength. Such constructions would carry over directly. One subtle point is how to address coding for the higher and very noisy levels. In fact, achieving reliability would be next to impossible at such low rates, and one would therefore not communicate over these levels, and just achieve channel resolvability using bits of private randomness. The top most levels have a rate that vanishes with the block length, and one concern is whether polarization happens fast enough at these levels. Lemma 6 shows that the rate decays exponentially with q and as the inverse of m. Polarization, however, only happens at a rate $\frac{1}{m^\gamma}$ for some $\gamma < 1$, which will therefore force us to overestimate the number of random bits to input. Fortunately, the number of levels is logarithmic in m, so that the rate of private randomness remains negligible.

4.4.1 Coding scheme using polar codes

In the following theorem, we show that we can achieve the covert capacity of a BI-DMC using polar codes.

Theorem 1. *Fix positive constants ζ and δ. For n large enough, there exist low-complexity polar coding schemes for each level of MLC-PPM scheme described in previous sections with covert rate at least $\frac{\sqrt{2}\mathbb{D}(P_1\|P_0)}{\sqrt{\chi_2(Q_1\|Q_0)}} - \zeta$, with probability of error at most ζ, and $\mathbb{D}(P_{\mathbf{Z}}\|Q_0^{\otimes n}) \leqslant \delta + \zeta$.*

Proof. Let m be the least power of two greater than $\left\lfloor \sqrt{\frac{\chi_2(Q_1\|Q_0)n}{2\delta}} \right\rfloor$ and $\ell \triangleq \left\lfloor \frac{n}{m} \right\rfloor$. We consider an MLC-PPM scheme with the number of levels $q = \log_2 m$ in which the transmission occurs in blocks of ℓ PPM symbols of order m. Since Willie's channel is degraded w.r.t. Bob's channel, for each level $i \in [\![1, q]\!]$, the channel $W^i(z_1, \cdots, z^{2^i}|x_i)$ defined in (4.28) is also degraded w.r.t. the channel $W^i(y_1, \cdots, y_{2^i}|x_i)$ defined in (4.6). Let $Q_{\widetilde{Z}^{(i)}}$

represent the distribution induced at the output of $W^i(z_1, \cdots, z_{2^i}|x_i)$ when the input to the channel is uniformly distributed. Let

$$R_i^Y \triangleq I(X_i; \widetilde{Y}|X^{i+1:q}) - \frac{A}{\ell^{\epsilon\kappa}}, \quad \text{and} \quad R_i^Z \triangleq I(X_i; \widetilde{Z}|X^{i+1:q}) + \frac{A}{\ell^{\epsilon\kappa}}.$$

Then, for some $\beta \in]0, \frac{1}{2}[$ and $\epsilon \in]0, 1-2\beta[$, by [42, Proposition 3, Lemma 1, and Lemma 2], there exist constants κ, A, and C and low-complexity polar codes $\mathcal{C}_i$ for each level $i \in [\![1, q]\!]$ with rate $R_i \triangleq \max(R_i^Y, R_i^Z)$ such that if ℓR_i bits are coded into a binary codeword of length $\ell > 2^C$ and transmitted over both $W^i(y_1, \cdots, y_{2^i}|x_i)$ and $W^i(z_1, \cdots, z_{2^i}|x_i)$, for the distribution $P_{\widetilde{\mathbf{Z}}^{(i)}}$ induced at the output of the channel $W^i(z_1, \cdots, z_{2^i}|x_i)$, we have $\mathbb{V}\Big(P_{\widetilde{\mathbf{Z}}^{(i)}}, Q_{\widetilde{Z}^{(i)}}^{\otimes\ell}\Big) \leqslant \mathcal{O}(\sqrt{\ell 2^{-\ell^\beta}})$ and a probability of error upper-bounded by $\mathcal{O}(\ell 2^{-\ell^\beta})$. Note that we need $R_i \geqslant R_i^Z$ to ensure channel resolvability and when $R_i^Z > R_i^Y$, we need key with rate $R_i^Z - R_i^Y$.

If the decoder uses the successive cancellation decoder of polar codes, by Lemma 10, the probability of error is upper bounded by $\mathcal{O}(q\ell 2^{-\ell^\beta})$, which is less than ζ for large enough n. For covertness, first notice that by Lemma 11, we have $\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes\ell}\big) \leqslant \mathcal{O}(q\sqrt{\ell 2^{-\ell^\beta}})$. Thus, from (4.29) and (4.30), we have

$$\begin{aligned}\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|Q_0^{\otimes n}\big) \leqslant \ell \log\left(\frac{1}{\min_{\widetilde{z}} Q_{\text{PPM}}^m(\widetilde{z})}\right) \mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes\ell}\big) + \mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes\ell}\|Q_0^{\otimes n}\big) \\ + 2\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes\ell}\big) \max_{\widetilde{z}} \ell \left|\log \frac{Q_{\text{PPM}}^m(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})}\right|. \quad (4.31)\end{aligned}$$

We have

$$\begin{aligned}Q_{\text{PPM}}^m(\widetilde{z}) &= \frac{1}{m}\sum_{i=1}^m Q_1(z_i) \prod_{j=1, j\neq i}^m Q_0(z_j) &(4.32)\\ &\geqslant \frac{1}{m}\mu_1\mu_0^{m-1}, &(4.33)\end{aligned}$$

where $\mu_0 = \min_{z \in \text{supp}(Q_0)} Q_0(z)$ and $\mu_1 = \min_{z \in \text{supp}(Q_1)} Q_1(z)$. Hence,

$$\log\left(\frac{1}{\min_{\widetilde{z}} Q_{\text{PPM}}^m(\widetilde{z})}\right) \leqslant (m-1)\log\frac{1}{\mu_0} + \log\frac{m}{\mu_1}. \tag{4.34}$$

We also have

$$\frac{Q_{\text{PPM}}^m(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})} = \frac{1}{m}\sum_{i=1}^{m}\frac{Q_1(z_i)}{Q_0(z_i)}. \tag{4.35}$$

Hence,

$$\frac{\mu_1}{m} \leqslant \frac{Q_{\text{PPM}}^m(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})} \leqslant \frac{1}{\mu_0}, \tag{4.36}$$

and

$$\left|\log\frac{Q_{\text{PPM}}^m(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})}\right| \leqslant \max\left(\log\frac{m}{\mu_1}, \log\frac{1}{\mu_0}\right). \tag{4.37}$$

For large enough m, we have

$$\left|\log\frac{Q_{\text{PPM}}^m(\widetilde{z})}{Q_0^{\otimes m}(\widetilde{z})}\right| \leqslant \log\frac{m}{\mu_1}. \tag{4.38}$$

From (3.18), we obtain

$$\mathbb{D}\big((Q_{\text{PPM}}^m)^{\otimes \ell} \| Q_0^{\otimes n}\big) \leqslant \delta + \mathcal{O}(\frac{1}{m}). \tag{4.39}$$

Therefore, (4.31) becomes

$$\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| Q_0^{\otimes n}\big) \leqslant \delta + \mathcal{O}(\frac{1}{m}) + \mathcal{O}(q\ell^2\sqrt{\ell 2^{-\ell^\beta}}), \tag{4.40}$$

which is less than $\delta + \zeta$ for large enough ℓ.

Finally, the number of transmitted bits is

$$\ell R_i^Y \geqslant \ell \sum_{i=1}^{q} \left(I(X_i; \widetilde{Y}|X^{i+1:q}) - \frac{A}{\ell^{\epsilon\kappa}} \right) \tag{4.41}$$

$$= \ell \left(I(\widetilde{X}; \widetilde{Y}) - \mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right) \right). \tag{4.42}$$

The covert throughput is given by

$$\frac{\ell \sum_{i=1}^{q} R_i^Y}{\sqrt{m\ell\delta}} \geqslant \sqrt{\frac{\ell}{m\delta}} \left(I(\widetilde{X}; \widetilde{Y}) - \mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right) \right) \tag{4.43}$$

$$= \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} \left(I(\widetilde{X}; \widetilde{Y}) - \mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right) \right) \tag{4.44}$$

$$= \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} \left(\mathbb{D}(P_1\|P_0) - \mathcal{O}\left(\frac{1}{m}\right) - \mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right) \right). \tag{4.45}$$

From Section 3.3.6, we know that for degraded case, $I(X_i; \widetilde{Y}|X^{i+1:q}) \geqslant I(X_i; \widetilde{Z}|X^{i+1:q})$ for all $i \in [\![1, q]\!]$. Hence, the number of key bits required is

$$\ell(R_i^Z - R_i^Y) \leqslant \ell \sum_{i=1}^{q} \frac{2A}{\ell^{\epsilon\kappa}} = \ell\mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right), \tag{4.46}$$

and the key throughput is

$$\frac{\ell(R_i^Z - R_i^Y)}{\sqrt{m\ell\delta}} \leqslant \mathcal{O}\left(\frac{q}{\ell^{\epsilon\kappa}}\right). \tag{4.47}$$

Hence, this coding scheme achieves the covert capacity. □

4.4.2 Coding scheme using invertible extractors

We can further reduce the complexity of the coding scheme using invertible extractors for channel resolvability. Since the capacity of the higher levels goes to zero exponentially, we can use the first few levels to code for reliability and only use the higher levels to achieve channel resolvability. Let u represent the number of levels used to code for reliability. We

can simplify the code construction by constructing a single code for all the higher levels. If we take the levels from $u+1$ to q as a single channel, the corresponding channel is given by

$$W^{u+1:q}\left(\widetilde{z}|x_{u+1:q}\right) = \frac{1}{2^u} \sum_{k \in \mathcal{A}^q(x_{u+1:q})} Q_0^{\otimes 2^q}(\widetilde{z}) \frac{Q_1(z_k)}{Q_0(z_k)}. \tag{4.48}$$

This channel is symmetric in the sense that for all $x_{u+1:q}$ and $x'_{u+1:q}$, there exists a permutation of the components of $\widetilde{z}$ denoted by $\pi_{x_{u+1:q}+x'_{u+1:q}}$ such that

$$W^{u+1:q}(\pi_{x_{u+1:q}+x'_{u+1:q}}(\widetilde{z})|x_{u+1:q}) = W^{u+1:q}(\widetilde{z}|x'_{u+1:q}). \tag{4.49}$$

We follow a construction of codes using invertible extractors similar to the ones used in [50, 40]. Let Ext be a two-universal extractor defined as

$$\text{Ext} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell} \to \mathbb{F}_2^{(q-u)\ell-k} : (s, x) \mapsto b,$$

and Inv be the inverter of Ext defined as

$$\text{Inv} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell-k} \times \mathbb{F}_2^k \to \mathbb{F}_2^{(q-u)\ell} : (s, b, r) \mapsto \mathbf{x}_{u+1:q}.$$

Let $\mathcal{P}_{s,b} \triangleq \{x \in \mathbb{F}_2^{(q-u)\ell} : \text{Ext}(s, x) = b\}$. We assume that Ext is regular, that is, $\left\{\mathcal{P}_{s,b}\right\}_{b \in \mathbb{F}_2^{(q-u)\ell-k}}$ forms a partition of $\mathbb{F}_2^{(q-u)\ell}$ into bins of equal size. For a given $s \in \mathbb{S}$ and $b \in \mathbb{F}_2^{(q-u)\ell-k}$, the encoder ϕ is given by

$$\phi : \mathbb{F}_2^k \to \mathbb{F}_2^{(q-u)\ell} : r \mapsto \text{Inv}(s, b, r).$$

Let $P_{\widetilde{\mathbf{Z}}}$ be the distribution induced by the coding scheme given by

$$P_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}) = \sum_{\mathbf{x}_{u+1:q} \in \mathcal{P}_{s,b}} \frac{1}{|\mathcal{P}_{s,b}|} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \tag{4.50}$$

$$= \sum_{\mathbf{x}_{u+1:q} \in \mathcal{P}_{s,b}} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}), \tag{4.51}$$

and $Q_{\widetilde{\mathbf{Z}}}$ be the distribution induced by an uniformly distributed input given by

$$Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}) = \sum_{\mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}). \tag{4.52}$$

Lemma 12. *The encoder ϕ defined above with rate $R_{u+1:q} = \frac{k}{\ell}$ with s and b selected randomly according to uniform distributions q_S and q_B, respectively, satisfies*

$$\lim_{\ell \to \infty} \mathbb{E}_{S,B} \left[\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| Q_{\widetilde{\mathbf{Z}}}\big) \right] = 0, \tag{4.53}$$

if $R_{u+1:q} > I(X_{u+1:q}; \widetilde{Z}) + 2\epsilon H(\widetilde{Z})$.

Proof. See Appendix 4.C. □

We now employ a specific extractor. For $s \in \mathbb{S} = \mathbb{F}_2^{(q-u)\ell} \backslash \{\mathbf{0}\}$, define

$$\text{Ext} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell} \to \mathbb{F}_2^{(q-u)\ell-k} : (s, x) \mapsto b \triangleq (s^{-1} \odot x)|_{[\![1,(q-u)\ell-k]\!]}, \tag{4.54}$$

where $\odot$ is the multiplication in the field $\mathbb{F}_2^{(q-u)\ell}$ and $(\cdot)|_{[\![1,(q-u)\ell-k]\!]}$ represents the bits in the positions $[\![1, (q-u)\ell]\!]$. Ext is a two-universal hash function [50], whose inverter is given by

$$\text{Inv} : \mathbb{S} \times \mathbb{F}_2^{(q-u)\ell-k} \times \mathbb{F}_2^k \to \mathbb{F}_2^{(q-u)\ell} : (s, b, r) \mapsto s \odot (b \| r), \tag{4.55}$$

where $(\cdot\|\cdot)$ denotes the concatenation of two sequences of bits. We now show that for the encoder implemented using this inverter, the divergence is same for any b. We can express

$Q_{\widetilde{\mathbf{Z}}}$ as follows:

$$Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}) = \sum_{\mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \tag{4.56}$$

$$= \sum_{b\in\mathbb{F}_2^{(q-u)\ell-k}} \sum_{r\in\mathbb{F}_2^k} \frac{1}{2^{(q-u)\ell}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|s\odot(b||r)) \tag{4.57}$$

$$= \sum_{b\in\mathbb{F}_2^{(q-u)\ell-k}} \sum_{r\in\mathbb{F}_2^k} \frac{1}{2^{(q-u)\ell}} W^{(u+1:q)^{\otimes \ell}}(\pi^{\ell}_{s\odot(b'||0)}(\widetilde{\mathbf{z}})|s\odot(b\oplus b'||r)) \tag{4.58}$$

$$= Q_{\widetilde{\mathbf{Z}}}(\pi^{\ell}_{s\odot(b'||0)}(\widetilde{\mathbf{z}})). \tag{4.59}$$

Taking expectation of the divergence between $P_{\widetilde{\mathbf{Z}}}$ and $Q_{\widetilde{\mathbf{Z}}}$ over S for a fixed b, we have

$$\mathbb{E}_{S,B=b}\left[\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|Q_{\widetilde{\mathbf{Z}}}\big)\right] = \sum_s q_S(s) \sum_{\widetilde{\mathbf{z}}} \sum_{\mathbf{x}_{u+1:q}\in\mathcal{P}_{s,b}} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q})$$
$$\log\left[\frac{\sum_{\bar{\mathbf{x}}_{u+1:q}\in\mathcal{P}_{s,b}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.60}$$

$$= \sum_s q_S(s) \sum_{\widetilde{\mathbf{z}}} \sum_{r\in\mathbb{F}_2^k} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|s\odot(b||r))$$
$$\log\left[\frac{\sum_{r'\in\mathbb{F}_2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|s\odot(b||r'))}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.61}$$

$$= \sum_s q_S(s) \sum_{\widetilde{\mathbf{z}}} \sum_{r\in\mathbb{F}_2^k} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\pi^{\ell}_{s\odot(b\oplus b'||0)}(\widetilde{\mathbf{z}})|s\odot(b'||r))$$
$$\log\left[\frac{\sum_{r'\in\mathbb{F}_2^k} W^{(u+1:q)^{\otimes \ell}}(\pi^{\ell}_{s\odot(b\oplus b'||0)}(\widetilde{\mathbf{z}})|s\odot(b'||r'))}{2^k Q_{\widetilde{\mathbf{Z}}}(\pi^{\ell}_{s\odot(b\oplus b'||0)}(\widetilde{\mathbf{z}}))}\right] \tag{4.62}$$

$$= \sum_s q_S(s) \sum_{\widetilde{\mathbf{z}}'} \sum_{r\in\mathbb{F}_2^k} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}'|s\odot(b'||r))$$
$$\log\left[\frac{\sum_{r'\in\mathbb{F}_2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}'|s\odot(b'||r'))}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}')}\right] \tag{4.63}$$

$$= \mathbb{E}_{S,B=b'}\left[\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}}\|Q_{\widetilde{\mathbf{Z}}}\big)\right]. \tag{4.64}$$

Hence, from (4.53)

$$\lim_{\ell\to\infty} \mathbb{E}_{S,B=b}\left[\mathbb{D}\big(P_{\tilde{\mathbf{Z}}}\|Q_{\tilde{\mathbf{Z}}}\big)\right] = \lim_{\ell\to\infty} \mathbb{E}_{S,B}\left[\mathbb{D}\big(P_{\tilde{\mathbf{Z}}}\|Q_{\tilde{\mathbf{Z}}}\big)\right] = 0. \tag{4.65}$$

From the above, we conclude that the choice of b is irrelevant for achieving channel resolvability. Hence, we can choose $b = 0$. Finally, note that we do not require S to be private, so we can use publicly available common randomness as the source for S, and we can reduce the rate of S by using a chaining strategy similar to the one in [40].

APPENDIX

4.A Proof of Lemma 10

Let $\mathcal{E}_i \triangleq \{\widehat{W}_i \neq W_i\}$. We can express the probability of error as follows:

$$\mathbb{P}\Big((\widehat{W}_1, \cdots, \widehat{W}_q) \neq (W_1, \cdots, W_q)\Big) = \mathbb{P}\left(\bigcup_{i=1}^{q} \mathcal{E}_i\right) \tag{4.66}$$

$$= \mathbb{P}\left(\bigcup_{i=1}^{q}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right)\right) \tag{4.67}$$

$$= \sum_{i=1}^{q} \mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right). \tag{4.68}$$

Expanding one term of the summation, we obtain

$$\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right) = \mathbb{P}\Big(\widehat{W}_i \neq W_i, \widehat{W}_{i+1} = W_{i+1}, \cdots, \widehat{W}_q = W_q\Big) \tag{4.69}$$

$$= \sum_{w_i} \cdots \sum_{w_q} \mathbb{P}\Big(\widehat{W}_i \neq w_i, W_i = w_i, \widehat{W}_{i+1} = W_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = W_q = w_q\Big) \tag{4.70}$$

$$= \sum_{w_i} \cdots \sum_{w_q} \mathbb{P}\Big(\widehat{W}_{i+1} = W_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = W_q = w_q\Big)\mathbb{P}(W_i = w_i)$$

$$\mathbb{P}\Big(\widehat{W}_i \neq w_i \,\Big|\, W_i = w_i, \widehat{W}_{i+1} = W_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = W_q = w_q\Big), \tag{4.71}$$

and

$$\mathbb{P}\Big(\widehat{W}_i \neq w_i | \widehat{W}_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = w_q, W_{i+1} = w_{i+1}, \ldots, W_q = w_q\Big)$$

$$= \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}}) \neq w_i} \mathbb{P}\Big(\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})} = \widetilde{\mathbf{y}} | \mathbf{X}_i = \mathbf{x}_i, \ldots, \mathbf{X}_q = \mathbf{x}_q\Big) \tag{4.72}$$

$$= \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1}) \prod_{j=1}^{\ell} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,\mathcal{A}^i(x_{1:i,j})}\right)}{P_0\left(y_{j,\mathcal{A}^i(x_{1:i,j})}\right)} \tag{4.73}$$

$$= \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1}) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \prod_{j=1}^{\ell} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,\mathcal{A}^i(x_{1:i,j})}\right)}{P_0\left(y_{j,\mathcal{A}^i(x_{1:i,j})}\right)} \tag{4.74}$$

$$= \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1}) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,k}\right)}{P_0\left(y_{j,k}\right)} \tag{4.75}$$

$$= \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,k}\right)}{P_0\left(y_{j,k}\right)}. \tag{4.76}$$

where (4.75) results from the assumption on the decoding region. In fact, each term inside $\sum_{k\in\mathcal{A}^i(x_{i,j})}$ can be obtained by permuting $\widetilde{y}_j$ by one of the permutation defined earlier. Since, all those permutations are included in the set $\{\widetilde{\mathbf{y}} : \psi_i(\widetilde{\mathbf{y}}) \neq w_i\}$, the same product term is repeated $\left(2^{i-1}\right)^{\ell}$ times.

Hence,

$$\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right)$$

$$= \sum_{w_i} \cdots \sum_{w_q} \mathbb{P}\left(\widehat{W}_{i+1} = W_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = W_q = w_q\right) \mathbb{P}(W_i = w_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,k}\right)}{P_0\left(y_{j,k}\right)} \tag{4.77}$$

$$= \sum_{w_i} \mathbb{P}(W_i = w_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1\left(y_{j,k}\right)}{P_0\left(y_{j,k}\right)} \sum_{w_{i+1}} \cdots \sum_{w_q} \mathbb{P}\left(\widehat{W}_{i+1} = W_{i+1} = w_{i+1}, \ldots, \widehat{W}_q = W_q = w_q\right) \tag{4.78}$$

$$\leqslant \sum_{w_i} \mathbb{P}(W_i = w_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(y_{j,k})}{P_0(y_{j,k})}$$
$$\sum_{w_{i+1}} \cdots \sum_{w_q} \mathbb{P}(W_{i+1} = w_{i+1}, \ldots, W_q = w_q) \quad (4.79)$$

$$= \sum_{w_i} \mathbb{P}(W_i = w_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(y_{j,k})}{P_0(y_{j,k})}. \quad (4.80)$$

For the equivalent channel, the probability of error for a given input message is

$$\mathbb{P}\Big(\widehat{W} \neq w_i | W_i = w_i\Big) = \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \prod_{j=1}^{\ell} \frac{1}{2^{i-1}} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(y_{j,k})}{P_0(y_{j,k})}. \quad (4.81)$$

Hence, the probability of error for the equivalent channel is given by

$$\mathbb{P}\Big(\widehat{W}_i \neq W_i\Big) = \sum_{w_i} \mathbb{P}(W_i = w_i) \mathbb{P}\Big(\widehat{W} \neq w_i | W_i = w_i\Big) \quad (4.82)$$

$$= \sum_{w_i} \mathbb{P}(W_i = w_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\widetilde{\mathbf{y}})\neq w_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(y_{j,k})}{P_0(y_{j,k})}. \quad (4.83)$$

Since we know that $\mathbb{P}\Big(\widehat{W}_i \neq W_i\Big) \leqslant \epsilon_i$, we have

$$\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right) \leqslant \mathbb{P}\Big(\widehat{W}_i \neq W_i\Big) \quad (4.84)$$

$$\leqslant \epsilon_i. \quad (4.85)$$

Therefore,

$$\mathbb{P}\Big((\widehat{W}_1, \cdots, \widehat{W}_q) \neq (W_1, \cdots, W_q)\Big) \leqslant \sum_{i=1}^{q} \epsilon_i, \quad (4.86)$$

which proves the lemma.

4.B Proof of Lemma 11

The result follows from observations similar to those in the proof of Lemma 10 regarding the symmetries of the PPM modulation. Specifically, consider level j. For all codewords $\mathbf{c}_{j+1}, \cdots, \mathbf{c}_q$ used in the upper levels, we have

$$\widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}}|\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{c}_j, \ldots, \mathbf{c}_q\right) = \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \frac{Q_1\big(z_{i,\mathcal{A}^q(x_{1:j-1,i}, c_{j:q,i})}\big)}{Q_0\big(z_{i,\mathcal{A}^q(x_{1:j-1,i}, c_{j:q,i})}\big)}, \tag{4.87}$$

and

$$\widetilde{W}^{\otimes \ell}\left(\widetilde{\mathbf{z}}|\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_j, \ldots, \mathbf{c}_q\right) = \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \frac{Q_1\big(z_{i,\mathcal{A}^q(x_{1:j,i}, c_{j+1:q,i})}\big)}{Q_0\big(z_{i,\mathcal{A}^q(x_{1:j,i}, c_{j+1:q,i})}\big)}. \tag{4.88}$$

Hence, we can bound $\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}^{(j-1)}, P_{\widetilde{\mathbf{Z}}}^{(j)}\right)$ by substituting the above equations in (4.27) as

$$\begin{aligned}
&\mathbb{V}\Big(P_{\widetilde{\mathbf{Z}}}^{(j-1)},\ P_{\widetilde{\mathbf{Z}}}^{j}\Big) \\
&\leqslant \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1} \in \mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}} \Bigg| \frac{1}{2^\ell} \sum_{\mathbf{x_1}} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_{j-1}}} \\
&\qquad \Bigg(\frac{1}{M_j} \sum_{\mathbf{c}_j \in \mathcal{C}_j} \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \frac{Q_1\big(z_{i,\mathcal{A}^q(x_{1:j-1,i}, c_{j:q,i})}\big)}{Q_0\big(z_{i,\mathcal{A}^q(x_{1:j-1,i}, c_{j:q,i})}\big)} \\
&\qquad\qquad - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \prod_{i=1}^{\ell} Q_0^{\otimes 2^q}(\widetilde{z}_i) \frac{Q_1\big(z_{i,\mathcal{A}^q(x_{1:j,i}, c_{j+1:q,i})}\big)}{Q_0\big(z_{i,\mathcal{A}^q(x_{1:j,i}, c_{j+1:q,i})}\big)} \Bigg) \Bigg|
\end{aligned} \tag{4.89}$$

$$\overset{(a)}{=} \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1}\in\mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q\in\mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}} \prod_{i'=1}^{\ell} \prod_{k'\in\mathcal{A}^q(c_{j+1:q,i'})^c} Q_0(z_{i',k'})$$

$$\sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})}} \left| \frac{1}{2^\ell} \sum_{\mathbf{x_1}} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_{j-1}}} \right.$$

$$\left(\frac{1}{M_j} \sum_{\mathbf{c}_j\in\mathcal{C}_j} \prod_{i=1}^{\ell} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,\mathcal{A}^q(x_{1:j-1,i},c_{j:q,i})})}{Q_0(z_{i,\mathcal{A}^q(x_{1:j-1,i},c_{j:q,i})})} \right.$$

$$\left. \left. - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \prod_{i=1}^{\ell} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,\mathcal{A}^q(x_{1:j,i},c_{j+1:q,i})})}{Q_0(z_{i,\mathcal{A}^q(x_{1:j,i},c_{j+1:q,i})})} \right) \right| \tag{4.90}$$

$$= \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1}\in\mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q\in\mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}} \prod_{i'=1}^{\ell} \prod_{k'\in\mathcal{A}^q(c_{j+1:q,i'})^c} Q_0(z_{i',k'})$$

$$\sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})}} \left| \frac{1}{M_j} \sum_{\mathbf{c}_j\in\mathcal{C}_j} \prod_{i=1}^{\ell} \frac{1}{2^{j-1}} \sum_{x_{1,i}} \cdots \sum_{x_{j-1,i}} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,\mathcal{A}^q(x_{1:j-1,i},c_{j:q,i})})}{Q_0(z_{i,\mathcal{A}^q(x_{1:j-1,i},c_{j:q,i})})} \right.$$

$$\left. - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \prod_{i=1}^{\ell} \frac{1}{2^{j-1}} \sum_{x_{1,i}} \cdots \sum_{x_{j-1,i}} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,\mathcal{A}^q(x_{1:j,i},c_{j+1:q,i})})}{Q_0(z_{i,\mathcal{A}^q(x_{1:j,i},c_{j+1:q,i})})} \right| \tag{4.91}$$

$$= \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1}\in\mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q\in\mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}} \prod_{i'=1}^{\ell} \prod_{k'\in\mathcal{A}^q(c_{j+1:q,i'})^c} Q_0(z_{i',k'})$$

$$\sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})}} \left| \frac{1}{M_j} \sum_{\mathbf{c}_j\in\mathcal{C}_j} \prod_{i=1}^{\ell} \frac{1}{2^{j-1}} \sum_{k\in\mathcal{A}^q(c_{j:q,i})} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,k})}{Q_0(z_{i,k})} \right.$$

$$\left. - \frac{1}{2^\ell} \sum_{\mathbf{x}_j} \prod_{i=1}^{\ell} \frac{1}{2^{j-1}} \sum_{k\in\mathcal{A}^q(x_{j,i},c_{j+1:q,i})} Q_0^{\otimes 2^j}(\widetilde{z}_{i,\mathcal{A}^q(c_{j+1:q,i})}) \frac{Q_1(z_{i,k})}{Q_0(z_{i,k})} \right| \tag{4.92}$$

$$= \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1}\in\mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q\in\mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}} \prod_{i'=1}^{\ell} \prod_{k'\in\mathcal{A}^q(c_{j+1:q,i'})^c} Q_0(z_{i',k'})$$

$$\sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})}} \left| P_{\widetilde{\mathbf{Z}}^{(j)}}(\widetilde{\mathbf{z}}^{(j)}(\mathbf{c}_{j+1:q})) - Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}(\widetilde{\mathbf{z}}^{(j)}(\mathbf{c}_{j+1:q})) \right| \tag{4.93}$$

$$= \frac{1}{M_{j+1}} \sum_{\mathbf{c}_{j+1} \in \mathcal{C}_{j+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}} \prod_{i'=1}^{\ell} \prod_{k' \in \mathcal{A}^q(c_{j+1:q,i'})^c} Q_0(z_{i',k'}) \mathbb{V}\big(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}\big) \tag{4.94}$$

$$= \mathbb{V}\big(P_{\widetilde{\mathbf{Z}}^{(j)}}, Q_{\widetilde{Z}^{(j)}}^{\otimes \ell}\big) \tag{4.95}$$

$$\leqslant \delta_j, \tag{4.96}$$

where (a) follows from dividing the summation over $\widetilde{\mathbf{z}}$ into summation over components $\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})}$ defined as in (4.20) and its complementary components denoted by $\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{c}_{j+1:q})^c}$.

Hence, from (4.25), we have

$$\mathbb{V}\left(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\right) \leqslant \sum_{j}^{q} \delta_j, \tag{4.97}$$

which proves the lemma.

4.C Proof of Lemma 12

We define a typical set as follows:

$$\mathcal{T}_\epsilon^\ell(X_{u+1:q}, \widetilde{Z}) \triangleq \Bigg\{ (\mathbf{x}_{u+1}, \ldots, \mathbf{x}_q, \widetilde{\mathbf{z}}) \in \mathcal{X}^\ell \times \cdots \times \mathcal{X}^\ell \times \widetilde{\mathcal{Z}}^\ell : \frac{1}{\ell} \log \frac{W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q})}{Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} < I(X_{u+1:q}; \widetilde{Z}) + \epsilon \Bigg\}. \tag{4.98}$$

We have

$$\mathbb{E}_{S,B}\left[\mathbb{D}\big(P_{\widetilde{\mathbf{Z}}} \| Q_{\widetilde{\mathbf{Z}}}\big)\right] = \sum_{s,b} q_S(s) q_B(b) \sum_{\widetilde{\mathbf{z}}} \sum_{\mathbf{x}_{u+1:q} \in \mathcal{P}_{s,b}} \frac{1}{2^k} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \log \left[\frac{\sum_{\bar{\mathbf{x}}_{u+1:q} \in \mathcal{P}_{s,b}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} \right] \tag{4.99}$$

$$\overset{(a)}{=} \sum_{s,b,\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q}} q_S(s)\frac{1}{2^{(q-u)\ell}}\mathbb{1}\left\{\text{Ext}(s,\mathbf{x}_{u+1:q})=b\right\}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \log\left[\frac{\sum_{\bar{\mathbf{x}}_{u+1:q}}\mathbb{1}\left\{\text{Ext}(s,\bar{\mathbf{x}}_{u+1:q})=b\right\}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.100}$$

$$\overset{(b)}{\leqslant} \sum_{\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \log\left[\frac{\sum_{s,\bar{\mathbf{x}}_{u+1:q}} q_S(s)\mathbb{1}\left\{\text{Ext}(s,\bar{\mathbf{x}}_{u+1:q})=\text{Ext}(s,\mathbf{x}_{u+1:q})\right\}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.101}$$

$$\overset{(c)}{\leqslant} \sum_{\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \log\left[\sum_{\bar{\mathbf{x}}_{u+1:q}\neq\mathbf{x}_{u+1:q}} W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q})\frac{2^{-((q-u)\ell-k)}}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})} + \frac{W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.102}$$

$$\overset{(d)}{\leqslant} \sum_{\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \log\left[1+\frac{W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q})}{2^k Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})}\right] \tag{4.103}$$

$$\begin{aligned}\overset{(e)}{\leqslant} &\sum_{(\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q})\in\mathcal{T}_\epsilon^\ell(X_{u+1:q},\widetilde{Z})} \frac{1}{2^{(q-u)\ell}}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q}) \\ &\qquad\times\log\left[1+2^{-\ell(R_{u+1:q}-I(X_{u+1:q};\widetilde{Z})-\epsilon)}\right] \\ &+\sum_{(\widetilde{\mathbf{z}},\mathbf{x}_{u+1:q})\notin\mathcal{T}_\epsilon^\ell(X_{u+1:q},\widetilde{Z})} \frac{1}{2^{(q-u)\ell}}W^{(u+1:q)^{\otimes\ell}}(\widetilde{\mathbf{z}}|\mathbf{x}_{u+1:q})\log\left[1+\mu_z^{-\ell}\right]\end{aligned} \tag{4.104}$$

$$\leqslant 2^{-\ell(R_{u+1:q}-I(X_{u+1:q};\widetilde{Z})-\epsilon)} + \frac{\ell}{\mu_z}\mathbb{P}\Big((\mathbf{X}_{u+1:q},\widetilde{\mathbf{Z}})\notin\mathcal{T}_\epsilon^\ell(X_{u+1:q},\widetilde{Z})\Big) \tag{4.105}$$

$$\leqslant 2^{-\ell(R_{u+1:q}-I(X_{u+1:q};\widetilde{Z})-\epsilon)} + \frac{\ell}{\mu_z}e^{-\frac{2\ell\epsilon^2}{q^2}} \overset{\ell\to\infty}{\longrightarrow} 0, \tag{4.106}$$

where

(a) follows from $q_B(b)=\frac{1}{2^{(q-u)\ell-k}}$ and from the definition of $\mathcal{P}_{s,b}$.

(b) follows by Jensen's inequality

(c) follows because Ext is two-universal and hence, for all $\bar{\mathbf{x}}_{u+1:q} \neq \mathbf{x}_{u+1:q}$, we have

$$\begin{aligned}\mathbb{P}_S(\text{Ext}(S, \bar{\mathbf{x}}_{u+1:q}) = \text{Ext}(S, \mathbf{x}_{u+1:q})) &= \sum_s q_S(s) \mathbb{1}\left\{\text{Ext}(s, \bar{\mathbf{x}}_{u+1:q}) = \text{Ext}(s, \mathbf{x}_{u+1:q})\right\} \\ &\leqslant 2^{-((q-u)\ell - k)},\end{aligned}$$

(d) follows because from the definition of $Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})$, we have

$$\sum_{\bar{\mathbf{x}}_{u+1:q} \neq \mathbf{x}_{u+1:q}} \frac{1}{2^{(q-u)\ell}} W^{(u+1:q)^{\otimes \ell}}(\widetilde{\mathbf{z}}|\bar{\mathbf{x}}_{u+1:q}) \leqslant Q_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}}).$$

(e) follows from bounding the term inside $\log$ from the definition of the typical set and using the definition $\mu_z = \min_{z \in \text{supp}(Q_Z)} Q_Z(z)$.

CHAPTER 5

CODES FOR COVERT COMMUNICATION OVER ADDITIVE WHITE GAUSSIAN CHANNELS

The information-theoretic analysis of covert communication for DMCs [16, 17] shows that the optimal signaling should be *sparse*, in that the fraction of "non-innocent symbols" in a codeword of length n should be $\mathcal{O}(\frac{1}{\sqrt{n}})$. From Section 2.6, we know that binary linear codes cannot achieve the covert capacity of BI-DMCs without shared secret key, which prompted the exploration of non-linear coding schemes. The proposed solution has been to rely on PPM, which was known to be optimal [44] with random non-binary code, and to circumvent the design of non-binary codes using MLC. This scheme systematically introduces sparsity in the inputs to the BI-DMC by coding over multiple levels and concentrating most of the information on a few levels. Most importantly, under MSD, the channels perceived at each level were shown to be *stationary*, thereby allowing the use of families of capacity- and resolvability-achieving codes.

In contrast, the information-theoretic analysis of covert communication for AWGN channels [16] shows that the optimal signaling should be *diffuse*, in that the covert capacity is achieved using Gaussian random codebooks whose variance scales with codeword length n as $\mathcal{O}(\frac{1}{\sqrt{n}})$. The covert capacity can also be achieved with binary phase-shift keying (BPSK)-modulation using appropriately scaled amplitude [51]. Unfortunately, this seemingly does not allow the application of the MLC-PPM scheme developed for DMCs. We, therefore, investigate the effectiveness of sparse modulation schemes for covert communication over AWGN channels and extend the MLC-PPM scheme to develop a low-complexity scheme. The rest of the paper is organized as follows. In Section 5.1, we introduce the exact problem setup. In Section 5.2, we study sparse modulation schemes for covert communication over AWGN channels, and in Section 5.3, we analyze a bi-

orthogonal PPM scheme [52] with two non-zero symbols for the "on" position. We briefly discuss the MLC-PPM scheme in Section 5.4.

5.1 Covert communication over AWGN channels

We consider a scenario in which a transmitter (Alice) communicates over an AWGN channel with transition probability $W_{Y|X}$ with a legitimate receiver (Bob) while avoiding detection from a warden (Willie) who observes Alice's transmission through another AWGN channel with transition probability $W_{Z|X}$. The two channels are described by

$$Y = X + N_b, \qquad Z = X + N_w, \tag{5.1}$$

where N_b and N_w have zero-mean Gaussian distributions with variances σ_b^2 and σ_w^2, respectively. The innocent input in the absence of communication is $X = 0$ and we denote the distributions $W_{Y|X=0}$ and $W_{Z|X=0}$ by P_0 and Q_0, respectively.

Alice encodes a message W uniformly drawn from a message set $\mathcal{M}$ to a codeword X^n of n real-valued symbols, which are observed by Bob and Willie as outputs of their channels Y^n and Z^n, respectively. Bob's estimate of the transmitted message is denoted by $\widehat{W}$. Let Q_Z^n be the distribution induced by the coding scheme and $Q_0^{\otimes n}$ be the innocent distribution induced by the all zero input. A code achieves covert throughput R with a covertness $\delta > 0$ if

$$\lim_{n\to\infty} \frac{\log|\mathcal{M}|}{\sqrt{n\delta}} \geqslant R, \qquad \lim_{n\to\infty} \mathbb{P}\left(\widehat{W} \neq W\right) = 0,$$
$$\text{and} \qquad \lim_{n\to\infty} \mathbb{D}(Q_Z^n \| Q_0^{\otimes n}) \leqslant \delta, \tag{5.2}$$

where $\mathbb{D}(\cdot\|\cdot)$ denotes the relative entropy between the distributions. The supremum of all achievable covert throughputs is called the covert capacity and denoted by L. For the case in which $\sigma_b = \sigma_w = \sigma$, the analysis in [16] shows that $L = 1$ nats. In our case, a similar

analysis shows that $L = \frac{\sigma_w^2}{\sigma_b^2}$ nats.

5.2 Sparse Modulation for Covert Communication

The analysis in [16] shows that the covert capacity of AWGN channels may be achieved with a Gaussian random code in which codewords are generated independent and identically distributed (i.i.d.) according to $\mathcal{N}(0, \rho_n)$ with $\rho_n \triangleq 2\sigma_w^2\sqrt{\delta/n}$. A refined analysis shows that random coding with uniformly distributed BPSK symbols and amplitude scaling as $\mathcal{O}(n^{-1/4})$ also achieves the covert capacity [51]. Unfortunately, such scaling results in a *diffuse* modulation for which coding schemes are challenging to design. The MLC-PPM scheme developed in the previous chapters for BI-DMC provides a method that uses independent codes over multiple levels and concentrates the communication rate on a few levels. The crucial component of that scheme is the use of PPM, a sparse modulation scheme. Motivated by the success of *sparse* modulation for code design, we study the effectiveness of sparse modulation over AWGN channels.

5.2.1 Sparse On-Off keying (OOK)

In an OOK scheme, one only uses two symbols as inputs to the channel: an "off" input 0 and an "on" input a. The input distribution on $\{0, a\}$ is denoted by Π_{α_n} with $\Pi_{\alpha_n}(a) = 1 - \Pi_{\alpha_n}(0) = \alpha_n$. The corresponding output distributions for Bob's and Willie's channels are P_Y and Q_Z, respectively. We define $P_{+a} \triangleq W_{Y|X=a}$, $Q_{+a} \triangleq W_{Z|X=a}$, and

$$P_Y(y) \triangleq (1-\alpha_n)P_0(y) + \alpha_n P_{+a}(y), \tag{5.3}$$

$$Q_Z(z) \triangleq (1-\alpha_n)Q_0(z) + \alpha_n Q_{+a}(z). \tag{5.4}$$

For distributions P_1, P_2 on $\mathcal{Z}$, we define for any integer $k \geqslant 1$

$$\chi_k(P_1||P_2) \triangleq \int_{\mathcal{Z}} \frac{(P_1(z) - P_2(z))^k}{P_2(z)^{k-1}} dz, \tag{5.5}$$

$$\eta_k(P_1||P_2) \triangleq \int_{\{z \in \mathcal{Z}: P_1(z) < P_2(z)\}} \frac{(P_1(z) - P_2(z))^k}{P_2(z)^{k-1}} dz. \tag{5.6}$$

By an analysis similar to [17], we can show that the mutual information between X and Y for the above distribution is

$$I(X;Y) = \mathbb{E}_{\Pi_{\alpha_n} W_{Y|X}} \left[\log \frac{W_{Y|X}(Y|X)}{P_Y(Y)} \right] \tag{5.7}$$

$$= \mathbb{E}_{\Pi_{\alpha_n} W_{Y|X}} \left[\log \frac{W_{Y|X}(Y|X)}{P_0(Y)} \right] - \mathbb{E}_{\Pi_{\alpha_n} W_{Y|X}} \left[\log \frac{P_Y(Y)}{P_0(Y)} \right] \tag{5.8}$$

$$= \alpha_n \mathbb{D}(P_{+a} \| P_0) - \mathbb{D}(P_Y \| P_0), \tag{5.9}$$

where

$$\mathbb{D}(P_{+a} \| P_0) = \mathbb{E}_{P_{+a}} \left[\log \frac{P_{+a}(Y)}{P_0(Y)} \right] \tag{5.10}$$

$$= \mathbb{E}_{P_{+a}} \left[\frac{aY}{\sigma_b^2} - \frac{a^2}{2\sigma_b^2} \right] = \frac{a^2}{2\sigma_b^2}. \tag{5.11}$$

One can also show that the relative entropy between distributions Q_Z and Q_0 can be bounded as follows for $\alpha_n < \frac{1}{2}$.

$$\mathbb{D}(Q_Z \| Q_0) \geqslant \frac{\alpha_n^2}{2} \chi_2(Q_{+a}||Q_0) - \alpha_n^3 \left(\frac{1}{2} \chi_3(Q_{+a}||Q_0) - \frac{2}{3} \eta_3(Q_{+a}||Q_0) \right) + \frac{2\alpha_n^4}{3} \eta_4(Q_{+a}||Q_0). \tag{5.12}$$

From the proof of [16, Theorem 5], the maximum covert throughput possible by re-

stricting the modulation scheme to OOK is given by

$$\widetilde{L} \leqslant \max_{\alpha_n, a} \liminf_{n\to\infty} \sqrt{\frac{n}{\delta}} I(X,Y), \text{ s.t. } \mathbb{D}(Q_Z \| Q_0) \leqslant \frac{\delta}{n}. \tag{5.13}$$

From (5.12), for $\mathbb{D}(Q_Z \| Q_0)$ to approach zero as n tends to infinity, we need α_n to tend to zero. Therefore, we have

$$\mathbb{D}(Q_Z \| Q_0) \geqslant \frac{\chi_2(Q_{+a}||Q_0)}{2} \alpha_n^2 + o(\alpha_n^2). \tag{5.14}$$

To satisfy the constraint in (5.13), α_n should satisfy

$$\alpha_n \leqslant \sqrt{\frac{2\delta}{n\chi_2(Q_{+a}||Q_0)}} + o(n^{-\frac{1}{2}}). \tag{5.15}$$

From (5.9) and (5.11), we have

$$I(X,Y) \leqslant \frac{\alpha_n a^2}{2\sigma_b^2} \leqslant \frac{a^2}{2\sigma_b^2} \sqrt{\frac{2\delta}{n\chi_2(Q_{+a}||Q_0)}} + o(n^{-\frac{1}{2}}). \tag{5.16}$$

Therefore,

$$\widetilde{L} \leqslant \frac{a^2}{\sigma_b^2 \sqrt{2\chi_2(Q_{+a}||Q_0)}} = \frac{a^2}{\sigma_b^2 \sqrt{2(\exp(\frac{a^2}{\sigma_w^2}) - 1)}}. \tag{5.17}$$

The maximum value of the above is approximately $0.57\frac{\sigma_w^2}{\sigma_b^2}$ for $a \approx \pm 1.26\sigma_w$, from which we conclude that coding over OOK cannot approach the covert capacity of AWGN channels.

5.2.2 Modified sparse OOK

We now show that by modifying the sparse OOK modulation scheme to allow both $-a$ and $+a$ symbols in the alphabet, we can achieve covert throughputs close to covert capacity. In

this modified OOK scheme, we use $\{0, -a, +a\}$ as the input alphabet with input distribution Π_{α_n} such that $\Pi_{\alpha_n}(-a) = \Pi_{\alpha_n}(+a) = \frac{\alpha_n}{2}$ and $\Pi_{\alpha_n}(0) = 1 - \alpha_n$. In this case, we have the output distributions P_Y and P_Z as follows:

$$P_{-a} \triangleq W_{Y|X=-a},\ P_{+a} \triangleq W_{Y|X=+a}, \tag{5.18}$$

$$Q_{-a} \triangleq W_{Z|X=-a},\ Q_{+a} \triangleq W_{Z|X=+a}, \tag{5.19}$$

$$P_{\pm a} \triangleq \frac{P_{-a} + P_{+a}}{2},\ Q_{\pm a} \triangleq \frac{Q_{-a} + Q_{+a}}{2}, \tag{5.20}$$

$$P_Y(y) \triangleq (1 - \alpha_n)P_0(y) + \alpha_n P_{\pm a}(y), \tag{5.21}$$

$$Q_Z(z) \triangleq (1 - \alpha_n)Q_0(z) + \alpha_n Q_{\pm a}(z). \tag{5.22}$$

By calculations similar to the OOK case, we get

$$I(X;Y) = \frac{\alpha_n}{2}\left(\mathbb{D}(P_{-a}\|P_0) + \mathbb{D}(P_{+a}\|P_0)\right) - \mathbb{D}(P_Y\|P_0) \tag{5.23}$$

and

$$\mathbb{D}(P_{-a}\|P_0) = \mathbb{D}(P_{+a}\|P_0) = \frac{a^2}{2\sigma_b^2}. \tag{5.24}$$

Also, using $\log(1 + x) \leqslant x - \frac{x^2}{2} + \frac{x^3}{3}$, we obtain

$$\mathbb{D}(Q_Z\|Q_0) \leqslant \frac{\alpha_n^2}{2}\chi_2(Q_{\pm a}\|Q_0) - \frac{\alpha_n^3}{6}\chi_3(Q_{\pm a}\|Q_0) + \frac{\alpha_n^4}{3}\chi_4(Q_{\pm a}\|Q_0). \tag{5.25}$$

Lemma 13. *For distributions $Q_{\pm a}$ and Q_0 defined above*

$$\chi_k(Q_{\pm a}||Q_0) = \sum_{\ell=0}^{k}\binom{k}{\ell}(-1)^{k-\ell}\frac{1}{2^\ell}\sum_{m=0}^{\ell}\binom{\ell}{m}\exp\left(\frac{[(\ell-2m)^2 - \ell]a^2}{2\sigma_w^2}\right). \tag{5.26}$$

Proof. We have

$$\chi_k(Q_{\pm a}||Q_0) = \int_{-\infty}^{\infty} \frac{(Q_{\pm a}(z) - Q_0(z))^k}{Q_0(z)^{k-1}} dz \tag{5.27}$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \int_{-\infty}^{\infty} \frac{Q_{\pm a}(z)^\ell}{Q_0(z)^{\ell-1}} dz. \tag{5.28}$$

We now expand the term

$$\int_{-\infty}^{\infty} \frac{Q_{\pm a}(z)^k}{Q_0(z)^{k-1}} dz = \int_{-\infty}^{\infty} \frac{1}{2^k} \frac{(Q_{+a}(z) + Q_{-a}(z))^k}{Q_0(z)^{k-1}} dz \tag{5.29}$$

$$= \frac{1}{2^k} \sum_{\ell=0}^{k} \binom{k}{\ell} \int_{-\infty}^{\infty} \frac{Q_{+a}(z)^\ell Q_{-a}(z)^{k-\ell}}{Q_0(z)^{k-1}} dz \tag{5.30}$$

$$= \frac{1}{2^k} \sum_{\ell=0}^{k} \binom{k}{\ell} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(-\frac{\ell(z-a)^2}{2\sigma_w^2} \right.$$
$$\left. -\frac{(k-\ell)(z+a)^2}{2\sigma_w^2} + \frac{(k-1)z^2}{2\sigma_w^2} \right) \tag{5.31}$$

$$= \frac{1}{2^k} \sum_{\ell=0}^{k} \binom{k}{\ell} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(\frac{-(z + (k-2\ell)a)^2}{2\sigma_w^2} \right.$$
$$\left. + \frac{[(k-2\ell)^2 - k]a^2}{2\sigma_w^2} \right) \tag{5.32}$$

$$= \frac{1}{2^k} \sum_{\ell=0}^{k} \binom{k}{\ell} \exp\left(\frac{[(k-2\ell)^2 - k]a^2}{2\sigma_w^2} \right). \tag{5.33}$$

Substituting the above equation in (5.28), we obtain (5.26). □

Using the series expansion of exponential in (5.26), we obtain

$$\chi_k(Q_{\pm a}||Q_0) = \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^\ell} \sum_{m=0}^{\ell} \binom{\ell}{m} \sum_{p}^{\infty} \frac{1}{p!} \left[\frac{(\ell-2m)^2 - \ell}{2} \right]^p \left(\frac{a^2}{\sigma_w^2} \right)^p \tag{5.34}$$

$$= \sum_{p}^{\infty} \frac{1}{p!} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+p}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[(\ell-2m)^2 - \ell \right]^p \left(\frac{a^2}{\sigma_w^2} \right)^p \tag{5.35}$$

$$= \sum_{p}^{\infty} c_{k,p} \left(\frac{a^2}{\sigma_w^2} \right)^p, \tag{5.36}$$

where

$$c_{k,p} = \frac{1}{p!} \sum_{\ell=0}^{k} \binom{k}{\ell} \frac{(-1)^{k-\ell}}{2^{\ell+p}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[(\ell-2m)^2 - \ell\right]^p .$$

One can show that $c_{k,0} = \delta[k-1],\ c_{k,1} = 0, c_{k,2} = \frac{1}{2}\delta[k-2]$, and $c_{k,3} = \delta[k-3]$.

$$c_{k,0} = \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^\ell} \sum_{m=0}^{\ell} \binom{\ell}{m} = \delta(k-1), \tag{5.37}$$

$$c_{k,1} = \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+1}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[(\ell-2m)^2 - \ell\right] \tag{5.38}$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+1}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[\ell^2 - 4m\ell + 4m^2 - \ell\right] \tag{5.39}$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+1}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[\ell(\ell-1) - 4(\ell-1)m + 4m(m-1)\right] \tag{5.40}$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+1}} \left[\ell(\ell-1)2^\ell - 4\ell(\ell-1)2^{\ell-1} + 4\ell(\ell-1)2^{\ell-2}\right] \tag{5.41}$$

$$= 0, \tag{5.42}$$

$$c_{k,2} = \frac{1}{2} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+2}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[(\ell-2m)^2 - \ell\right]^2 \tag{5.43}$$

$$= \frac{1}{2} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+2}} \ell(\ell-1)2^{\ell+1} \tag{5.44}$$

$$= \frac{1}{4} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \ell(\ell-1) \tag{5.45}$$

$$= \frac{1}{4} k(k-1) \sum_{\ell=2}^{k} \binom{k-2}{\ell-2} (-1)^{k-\ell} \tag{5.46}$$

$$= \frac{1}{2}\delta(k-2), \tag{5.47}$$

$$c_{k,3} = \frac{1}{6} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+3}} \sum_{m=0}^{\ell} \binom{\ell}{m} \left[(\ell-2m)^2 - \ell\right]^3 \tag{5.48}$$

$$= \frac{1}{6} \sum_{\ell=0}^{k} \binom{k}{\ell} (-1)^{k-\ell} \frac{1}{2^{\ell+3}} \ell(\ell-1)(\ell-2)2^{\ell+3} \tag{5.49}$$

$$= \frac{1}{6}k(k-1)(k-2)\sum_{\ell=3}^{k}\binom{k}{\ell}(-1)^{k-\ell} \tag{5.50}$$

$$= \delta(k-3). \tag{5.51}$$

Hence, from (5.25), we have

$$\mathbb{D}(Q_Z\|Q_0) \leqslant \frac{\alpha_n^2 a^4}{4\sigma_w^4} - \frac{\alpha_n^3 a^6}{6\sigma_w^6} + \sum_{p=4}^{\infty}\left(\frac{c_{2,p}}{2}\alpha_n^2 + \frac{c_{3,p}}{6}\alpha_n^3 + \frac{c_{4,p}}{3}\alpha_n^4\right)\left(\frac{a^2}{\sigma_w^2}\right)^p. \tag{5.52}$$

For a codebook of $|\mathcal{M}|$ codewords, each of length n generated i.i.d. with distribution Π_{α_n}, we have

$$I(X^n;Y^n) = nI(X;Y) = n\left(\frac{\alpha_n a^2}{2\sigma_b^2} - \mathbb{D}(P_Y\|P_0)\right), \tag{5.53}$$

$$\mathbb{D}(Q_Z^n\|Q_0^{\otimes n}) = n\mathbb{D}(Q_Z\|Q_0). \tag{5.54}$$

By choosing $\alpha_n = \sqrt{\frac{2\delta}{n\chi_2(Q_{\pm a}\|Q_0)}}$ for a given $\delta > 0$, we obtain

$$\mathbb{D}(Q_Z^n\|Q_0^{\otimes n}) \leqslant \delta + \mathcal{O}(n^{-1/2}), \tag{5.55}$$

$$\frac{I(X^n;Y^n)}{\sqrt{n\delta}} \geqslant \frac{a^2}{\sigma_b^2\sqrt{2\chi_2(Q_{\pm a}\|Q_0)}} - \mathcal{O}(n^{-1/2}) \tag{5.56}$$

$$= \frac{a^2}{\sigma_b^2\sqrt{2\left(\cosh(\frac{a^2}{\sigma_w^2}) - 1\right)}} - \mathcal{O}(n^{-1/2}). \tag{5.57}$$

Note that

$$\lim_{a\to 0}\frac{a^2}{\sigma_b^2\sqrt{2\left(\cosh(\frac{a^2}{\sigma_w^2}) - 1\right)}} = \frac{\sigma_w^2}{\sigma_b^2}. \tag{5.58}$$

This shows that we can achieve a covert throughput close to the optimal throughput by choosing a small enough and α_n as above. This results in a sparse modulation, which

allows us to implement an MLC-PPM scheme.

Remark 1. *With $\alpha_n = \alpha$ and $a = \sigma_w\sqrt{\frac{2}{\alpha}}\left(\frac{\delta}{n}\right)^{1/4}$, we have from* (5.52)*,*

$$\mathbb{D}(Q_Z^n \| Q_0^{\otimes n}) \leqslant \delta + \mathcal{O}(n^{-1/2}), \tag{5.59}$$

$$\frac{I(X^n;Y^n)}{\sqrt{n\delta}} \geqslant \frac{\sigma_w^2}{\sigma_b^2} - \mathcal{O}(n^{-1/2}). \tag{5.60}$$

This corresponds to a diffuse modulation scheme with uniformly distributed BPSK symbols that we mentioned in the beginning of this section [51].

5.3 Bi-orthogonal PPM for AWGN channel

In the spirit of [44] for DMCs, we now show how to replace the modified OOK scheme with a bi-orthogonal PPM to achieve the optimal throughput. We define the bi-orthogonal PPM alphabet of order m as $\widetilde{\mathcal{X}} = \{\widetilde{x}_{i,j}\}_{i\in[\![1,m]\!],j\in\{0,1\}}$, where $\widetilde{x}_{i,j}$ is vector of length m whose i-th element is $(-1)^j a$ and all other elements are 0. By selecting symbols uniformly from $\widetilde{\mathcal{X}}$ and transmitting them through m independent uses of the channels $W_{Y|X}$ and $W_{Z|X}$, denoted by the PPM super channels $\widetilde{W}_{\widetilde{Y}|\widetilde{X}}$ and $\widetilde{W}_{\widetilde{Z}|\widetilde{X}}$, we obtain the output distributions $P_{\widetilde{Y}}$ and $Q_{\widetilde{Z}}$, respectively. Let $\widetilde{X}$ represent the input random variable uniformly distributed on $\widetilde{\mathcal{X}}$ and $\widetilde{Y}$ and $\widetilde{Z}$ represent the corresponding output random variables.

The mutual information between $\widetilde{X}$ and $\widetilde{Y}$ is given by

$$\begin{aligned}
I(\widetilde{X};\widetilde{Y}) &= \mathbb{E}_{\frac{1}{2m}W_{Y|X}^{\otimes m}}\left[\log\frac{W_{Y|X}^{\otimes m}(\widetilde{Y}|\widetilde{X})}{P_{\widetilde{Y}}(\widetilde{Y})}\right] \\
&= \mathbb{E}_{\frac{1}{2m}W_{Y|X}^{\otimes m}}\left[\log\frac{W_{Y|X}^{\otimes m}(\widetilde{Y}|\widetilde{X})}{P_0^{\otimes m}(\widetilde{Y})}\right] - \mathbb{E}_{\frac{1}{2m}W_{Y|X}^{\otimes m}}\left[\frac{P_{\widetilde{Y}}(\widetilde{Y})}{P_0^{\otimes m}(\widetilde{Y})}\right] \\
&= \int_{\widetilde{\mathcal{Y}}}\sum_{i=1}^{m}\sum_{j\in\{0,1\}}\frac{1}{2m}W_{Y|X}^{\otimes m}(\widetilde{y}|\widetilde{x}_{i,j})\log\frac{P_{(-1)^j a}(y_i)}{P_0(y_i)}d\widetilde{y} - \mathbb{D}\big(P_{\widetilde{Y}}\|P_0^{\otimes m}\big) \qquad (5.61) \\
&= \sum_{i=1}^{m}\sum_{j\in\{0,1\}}\frac{1}{2m}\int_{\mathcal{Y}}P_{(-1)^j a}(y_i)\log\frac{P_{(-1)^j a}(y_i)}{P_0(y_i)}dy_i - \mathbb{D}\big(P_{\widetilde{Y}}\|P_0^{\otimes m}\big) \qquad (5.62)
\end{aligned}$$

$$= \sum_{i=1}^{m} \sum_{j\in\{0,1\}} \frac{1}{2m} \mathbb{D}\big(P_{(-1)^j a} \| P_0\big) - \mathbb{D}\big(P_{\widetilde{Y}} \| P_0^{\otimes m}\big) \tag{5.63}$$

$$= \frac{1}{2}\left(\mathbb{D}(P_{-a}\|P_0) + \mathbb{D}(P_{+a}\|P_0)\right) - \mathbb{D}\big(P_{\widetilde{Y}} \| P_0^{\otimes m}\big). \tag{5.64}$$

We now define

$$A_{\pm a}(z) \triangleq \frac{Q_{\pm a}(z) - Q_0(z)}{Q_0(z)}. \tag{5.65}$$

The relative entropy between $Q_{\widetilde{Z}}$ and $Q_0^{\otimes m}$ is given by

$$\mathbb{D}\big(Q_{\widetilde{Z}} \| Q_0^{\otimes m}\big) = \int_{\widetilde{\mathcal{Z}}} \sum_{i=1}^{m} \sum_{j\in\{0,1\}} \frac{1}{2m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{(-1)^j a}(z_i)}{Q_0(z_i)} \times \log\left(1 + \sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'})\right) d\widetilde{z} \tag{5.66}$$

$$= \int_{\widetilde{\mathcal{Z}}} \sum_{i=1}^{m} \frac{1}{m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{\pm a}(z_i)}{Q_0(z_i)} \log\left(1 + \sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'})\right) d\widetilde{z}. \tag{5.67}$$

We bound the above using the inequality $\log(1+x) \leqslant x - \frac{x^2}{2} + \frac{x^3}{3}$. First, note that

$$\int_{\mathcal{Z}} Q_0(z)(A_{\pm a}(z))^k dz = \chi_k(Q_{\pm a}||Q_0). \tag{5.68}$$

And

$$\begin{aligned}\int_{\mathcal{Z}} Q_{\pm a}(z)(A_{\pm a}(z))^k dz &= \int_{\mathcal{Z}} Q_0(z)(A_{\pm a}(z))^k dz + \int_{\mathcal{Z}} (Q_{\pm a}(z) - Q_0(z))(A_{\pm a}(z))^k dz \\ &= \int_{\mathcal{Z}} Q_0(z)(A_{\pm a}(z))^k + \int_{\mathcal{Z}} Q_0(z)(A_{\pm a}(z))^{k+1} \\ &= \chi_k(Q_{\pm a}||Q_0) + \chi_{k+1}(Q_{\pm a}||Q_0). \end{aligned} \tag{5.69}$$

We have

$$\int_{\widetilde{\mathcal{Z}}} \sum_{i=1}^{m} \frac{1}{m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{\pm a}(z_i)}{Q_0(z_i)} \sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'}) d\widetilde{z}$$

$$= \frac{1}{m^2} \sum_{i=1}^{m} \left(\int_{\mathcal{Z}} Q_{\pm a}(z_i) A_{\pm a}(z_i) dz_i + \sum_{i' \neq i} \int_{\mathcal{Z}} Q_0(z_{i'}) A_{\pm a}(z_{i'}) dz_{i'} \right) \tag{5.70}$$

$$= \frac{1}{m} \chi_2(Q_{\pm a} || Q_0). \tag{5.71}$$

Also,

$$\int_{\widetilde{z}} \sum_{i=1}^{m} \frac{1}{m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{\pm a}(z_i)}{Q_0(z_i)} \left(\sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'}) \right)^2$$

$$= \frac{1}{m^3} \sum_{i=1}^{m} \Bigg[\sum_{i' \neq i} \Bigg(\sum_{\substack{i'' \neq i \\ i'' \neq i'}} \int_{z_{i'}} Q_0(z_{i'}) A_{\pm a}(z_{i'}) \int_{z_{i''}} Q_0(z_{i''}) A_{\pm a}(z_{i''})$$

$$+ \sum_{i''=i'} \int_{z_{i'}} Q_0(z_{i'}) (A_{\pm a}(z_{i'}))^2$$

$$+ \int_{z_i} Q_{\pm a}(z_i) A_{\pm a}(z_i) \int_{z_{i'}} Q_0(z_{i'}) A_{\pm a}(z_{i'}) \Bigg)$$

$$+ \sum_{i'=i} \Bigg(\sum_{i'' \neq i} \int_{z_i} Q_{\pm a}(z_i) A_{\pm a}(z_i) \int_{z_{i''}} Q_0(z_{i''}) A_{\pm a}(z_{i''})$$

$$+ \sum_{i''=i} \int_{z_i} Q_{\pm a}(z_i) (A_{\pm a}(z_i))^2 \Bigg) \Bigg] \tag{5.72}$$

$$= \frac{1}{m^3} \sum_{i=1}^{m} [(m-1)\chi_2 + \chi_2 + \chi_3] \tag{5.73}$$

$$= \frac{1}{m} \chi_2(Q_{\pm a} || Q_0) + \frac{1}{m^2} \chi_3(Q_{\pm a} || Q_0). \tag{5.74}$$

$$\int_{\widetilde{z}} \sum_{i=1}^{m} \frac{1}{m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{\pm a}(z_i)}{Q_0(z_i)} \left(\sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'}) \right)^3$$

$$= \int_{\widetilde{z}} \sum_{i=1}^{m} \frac{1}{m} Q_0^{\otimes m}(\widetilde{z}) \frac{Q_{\pm a}(z_i)}{Q_0(z_i)} \sum_{i'=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'}) \sum_{i''=1}^{m} \frac{1}{m} A_{\pm a}(z_{i''}) \sum_{i'''=1}^{m} \frac{1}{m} A_{\pm a}(z_{i'''}) \quad (5.75)$$

$$= \frac{1}{m^2} \left(\chi_3(Q_{\pm a} || Q_0) + 3\chi_2(Q_{\pm a} || Q_0)^2 \right) + \frac{1}{m^3} \left(\chi_4(Q_{\pm a} || Q_0) - 3\chi_2(Q_{\pm a} || Q_0)^2 \right) \quad (5.76)$$

By similarly evaluating degree 2 and 3 terms, we obtain

$$\mathbb{D}\big(Q_{\widetilde{Z}} \| Q_0^{\otimes m}\big) \leqslant \frac{1}{2m} \chi_2(Q_{\pm a} || Q_0) + \frac{1}{m^2} \left(\chi_2(Q_{\pm a} || Q_0)^2 - \frac{1}{6} \chi_3(Q_{\pm a} || Q_0) \right) + \frac{1}{m^3} \left(\frac{1}{3} \chi_4(Q_{\pm a} || Q_0) - \chi_2(Q_{\pm a} || Q_0)^2 \right). \quad (5.77)$$

By an argument similar to the modified OOK case, one can show by random coding over $\ell = \lceil \frac{2\delta}{\chi_2(Q_{\pm a} || Q_0)} m \rceil$ super channel uses that this scheme achieves the covert capacity.

5.4 Multilevel coding scheme for AWGN channels

In this section, we provide an overview of the MLC-PPM coding scheme for covert communication over AWGN channels. For simplicity, we assume that $\sigma_w > \sigma_b$, i.e., Willie's

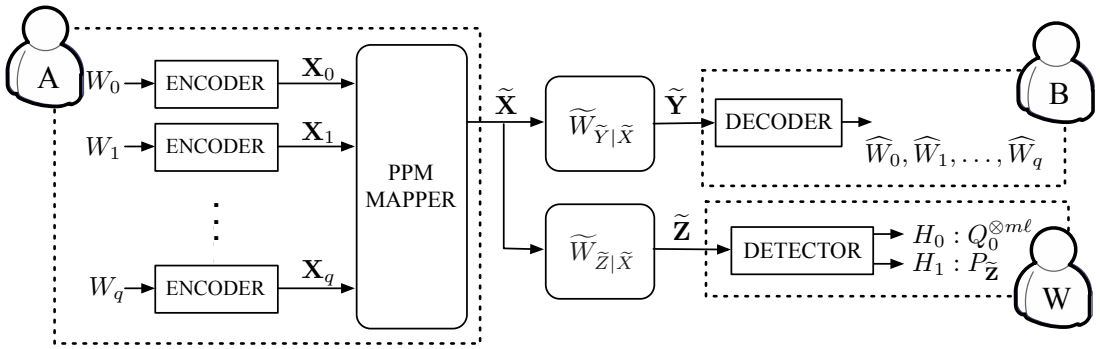

Figure 5.1: Setup for MLC-PPM scheme for AWGN channels.

channel is degraded w.r.t. Bob's channel, and that $q = \log_2 m$ is an integer, where m is the order of the bi-orthogonal PPM symbol. The MLC coding scheme is similar to the one introduced in Chapter 3 for DMC but requires $q+1$ levels to account for the choice between $-a$ and a at the position of non-zero symbol. As shown in Fig. 5.1, the input to i-th level for $i \in [\![0, q]\!]$ by W_i, with level 0 coding for the sign of the non-zero in the PPM symbol. The encoder for the i-th level selects the codeword indexed by W_i from the codebook $\mathcal{C}_i$ as the input $\mathbf{X}_i$ to the PPM mapper. The PPM mapper maps the inputs $X_{0:q}$ to $\widetilde{x}(X_{0:q})$, where the mapping $\widetilde{x} : \mathbb{F}_2^{q+1} \to \widetilde{X}^q$ is given by

$$\widetilde{x}(x_{0:q}) = \widetilde{x}_{\mathcal{A}^q(x_{1:q}),x_0} = \widetilde{x}_{d(x_{1:q}),x_0}, \tag{5.78}$$

where $d(x_{1:q})$ and $\mathcal{A}^q(x_{\mathcal{S}})$ are as defined in Section 3.1.

We decompose the mutual information between the input and the output as

$$I(\widetilde{X}; \widetilde{Y}) = I(X_{0:q}; \widetilde{Y}) = \sum_{i=0}^{q} I(X_i; \widetilde{Y}|X_{i+1:q}). \tag{5.79}$$

As identified in Section 3.3.4, this decomposition suggests an operating point for MLC for which the rate of each level is given by

$$R_i = I(X_i; \widetilde{Y}|X_{i+1:q}) - \frac{\varepsilon}{q+1}. \tag{5.80}$$

We also define the equivalent channel corresponding to each level as follows. The equivalent channel for 0-th level is

$$W^0(y|x_0) = P_{(-1)^{x_0}a}(y), \tag{5.81}$$

and the equivalent channel for i-th level for $i \in [\![1, q]\!]$ is

$$W^i(y_1, \ldots, y_{2^i}|x_i) = \frac{1}{2^i} \sum_{x_0 \in \mathbb{F}_2} \sum_{\mathcal{A}^i(x_i)} P_0^{2^i}(y_1, \ldots, y_{2^i}) \frac{P_{(-1)^{x_0} a}(y_{\mathcal{A}^i(x_{1:i})})}{P_0(y_{\mathcal{A}^i(x_{1:i})})} \tag{5.82}$$

$$= \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_{\pm a}(y_k)}{P_0(y_k)}. \tag{5.83}$$

Lemma 14. *For $i \in [\![0, q]\!]$, suppose (ϕ_i, ψ_i) is an encoder/decoder for the stationary channel W^i defined in* (5.81)-(5.83) *with rate R_i and probability of error ϵ_i, such that ψ_i satisfies the mild assumption on the decoding region defined in Section 4.1. Then, the overall MLC-PPM scheme has rate $R = \sum_{i=0}^{q} R_i$ and probability of error $\epsilon = \sum_{i=0}^{q} \epsilon_i$.*

Proof. The proof follows steps similar to those of Lemma 10, adapted to continuous output alphabets. □

Consider now the variational distance $\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big)$, where $P_{\widetilde{\mathbf{Z}}}$ denotes the distribution induced at the output of the PPM super-channel when coding over ℓ PPM symbols of order m, Q_{PPM}^m denotes the output distribution induced by a uniform input distribution on PPM symbols of order m, and $(Q_{\text{PPM}}^m)^{\otimes \ell}$ denotes product distribution over the ℓ uses of the super channel. For $i \in [\![0, q]\!]$, assume that the codebook for level i consists of M_i codewords $\mathcal{C}_i = \{\mathbf{c}(j_i)\}_{j_i=1}^{M_i}$. For $i \in [\![0, q]\!]$, we define

$$P_{\widetilde{\mathbf{Z}}}^{(i)}(\widetilde{\mathbf{z}}) = \frac{1}{2^\ell} \sum_{\mathbf{x_0} \in \{\mathbf{0},\mathbf{1}\}^\ell} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_{i-1}} \in \{\mathbf{0},\mathbf{1}\}^\ell} \frac{1}{M_i} \sum_{\mathbf{c}_i \in \mathcal{C}_i} \cdots \frac{1}{M_q}$$

$$\times \sum_{\mathbf{c}_q \in \mathcal{C}_q} \widetilde{W}^{\otimes \ell}(\widetilde{\mathbf{z}}|\mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{c}_i, \ldots, \mathbf{c}_q). \tag{5.84}$$

Note that $P_{\widetilde{\mathbf{Z}}}^{(0)}(\widetilde{\mathbf{z}}) = P_{\widetilde{\mathbf{Z}}}(\widetilde{\mathbf{z}})$ and $P_{\widetilde{\mathbf{Z}}}^{(q+1)}(\widetilde{\mathbf{z}}) = (Q_{\text{PPM}}^m)^{\otimes \ell}$. Using triangle inequality, we obtain

$$\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q_{\text{PPM}}^m)^{\otimes \ell}\big) \leqslant \sum_{i=0}^{q} \mathbb{V}\Big(P_{\widetilde{\mathbf{Z}}}^{(i)}, P_{\widetilde{\mathbf{Z}}}^{(i+1)}\Big). \tag{5.85}$$

We upper bound $\mathbb{V}\Big(P^{(i)}_{\widetilde{\mathbf{Z}}}, P^{(i+1)}_{\widetilde{\mathbf{Z}}}\Big)$ as

$$\begin{aligned}
\mathbb{V}\left(P^{(i)}_{\widetilde{\mathbf{Z}}}, P^{i+1}_{\widetilde{\mathbf{Z}}}\right) = \int_{\widetilde{\mathbf{z}}} \Bigg| & \frac{1}{2^\ell} \sum_{\mathbf{x_0}} \cdots \frac{1}{2^\ell} \sum_{\mathbf{x_{i-1}}} \frac{1}{M_{i+1}} \sum_{\mathbf{c}_{i+1} \in \mathcal{C}_{i+1}} \cdots \\
& \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \Bigg(\frac{1}{M_i} \sum_{\mathbf{c}_i \in \mathcal{C}_i} \widetilde{W}^{\otimes \ell} \left(\widetilde{\mathbf{z}} | \mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{c}_i, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_q\right) \\
& \qquad - \frac{1}{2^\ell} \sum_{\mathbf{x}_i} \widetilde{W}^{\otimes \ell} \left(\widetilde{\mathbf{z}} | \mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_q\right) \Bigg) \Bigg| \qquad (5.86) \\
\leqslant \frac{1}{M_{i+1}} & \sum_{\mathbf{c}_{i+1} \in \mathcal{C}_{i+1}} \cdots \frac{1}{M_q} \sum_{\mathbf{c}_q \in \mathcal{C}_q} \int_{\widetilde{\mathbf{z}}} \Bigg| \frac{1}{2^\ell} \sum_{\mathbf{x_0}} \cdots \frac{1}{2^\ell} \\
& \sum_{\mathbf{x_{i-1}}} \Bigg(\frac{1}{M_i} \sum_{\mathbf{c}_i \in \mathcal{C}_i} \widetilde{W}^{\otimes \ell} \left(\widetilde{\mathbf{z}} | \mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{c}_i, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_q\right) \\
& \qquad - \frac{1}{2^\ell} \sum_{\mathbf{x}_i} \widetilde{W}^{\otimes \ell} \left(\widetilde{\mathbf{z}} | \mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{c}_{i+1}, \ldots, \mathbf{c}_q\right) \Bigg) \Bigg| \qquad (5.87)
\end{aligned}$$

Then, the following holds for resolvability.

Lemma 15. *Consider the stationary channels*

$$\begin{aligned}
W^0(z|x_0) &= Q_{(-1)^{x_0} a}(z), \qquad (5.88) \\
W^i(z_1, \ldots, z_{2^i} | x_i) &= \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} Q_0^{\otimes 2^i}(z_1, \ldots, z_{2^i}) \\
&\quad \times \frac{Q_{\pm a}(z_k)}{Q_0(z_k)} \qquad \textit{for } i \in [\![1, q]\!]. \qquad (5.89)
\end{aligned}$$

For $i \in [\![0, q]\!]$, let $P_{\widetilde{\mathbf{Z}}^{(i)}}$ denote the output distribution induced by a code over this channel, and let $Q^{\otimes \ell}_{\widetilde{Z}^{(i)}}$ denote the product output distribution when the input is uniform. If $\mathbb{V}\Big(P_{\widetilde{\mathbf{Z}}^{(i)}}, Q^{\otimes \ell}_{\widetilde{Z}^{(i)}}\Big) \leqslant \delta_i$, then the same code ensures that $\mathbb{V}\Big(P^{(i)}_{\widetilde{\mathbf{Z}}}, P^{(i+1)}_{\widetilde{\mathbf{Z}}}\Big) \leqslant \delta_i$ irrespective of the code used for the higher levels. Moreover, when combined, these codes together ensure $\mathbb{V}\big(P_{\widetilde{\mathbf{Z}}}, (Q^m_{PPM})^{\otimes \ell}\big) \leqslant \sum_{i=0}^q \delta_i$.

Proof. The proof follows steps similar to those of Lemma 11, adapted to the continuous

output alphabets. □

Consequently, one can implement codes at each level such as those outlined in [53]. Moreover, as shown in Section 4.1, most of the information concentrates in the lower levels implying an efficient design by using higher levels just for resolvability.

CHAPTER 6

FORWARD RECONCILIATION FOR COVERT KEY GENERATION

Information reconciliation is an important step in the secret key generation. In secret key generation problem, we consider two legitimate users and one adversary. The objective of the legitimate users is to generate a secret key from their correlated observations and the public information exchanged between them. In the first step, both users agree on a common sequence by exchanging messages over a public channel. This step is known as the information reconciliation step. Then, they generate the secret key by applying a publicly agreed deterministic function on that common sequence. There are mainly two kinds of secret key generation models: source model and channel model. In the source model, two legitimate users and the adversary observe three correlated sequences. In the channel model, one of the legitimate users generate a sequence and the other user and the adversary observes the output of their respective channels. By restricting the exchange of public messages in the channel model problem to only one direction, we can classify the reconciliation protocol to two types: forward reconciliation in which the public messages are allowed only in the direction of channel and reverse reconciliation in which the public messages are allowed only in the reverse direction.

Along with the development in the classical covert communication, there have been several efforts to extend the ideas of covert communication to the quantum regime [54, 29, 55] and in particular to quantum key distribution [28, 56]. The most recent efforts in this direction have clarified the role of public communication, and clearly defined how information reconciliation impacts covertness for key generation over quantum channels [57]. While the introduction of covert constraints does not change the high-level principle of reconciliation [58], a subtle requirement is that the reconciliation message should remain independent of the eavesdropper's observations. While this does not pose any particular

difficulty from an information-theoretic perspective, this complicates the design of actual codes.

The objective of this chapter is to offer an initial low-complexity solution to the problem of forward reconciliation for covert key generation. The ideas are largely building on previous work by exploiting MLC and PPM, but reconciliation introduces specific subtleties that need to be properly addressed. For ease of exposition, the adversary is assumed classical but the results generalize to quantum adversaries.

6.1 Covert forward-reconciliation

We analyze covert forward-reconciliation for the channel model illustrated in Figure 6.1. Alice generates a sequence $\mathbf{X}$ of length n using a distribution Q_X under her control. Bob and Willie observe the sequence through BI-DMCs $(\mathcal{X}, W_{Y|X}, \mathcal{Y})$ and $(\mathcal{X}, W_{Z|X}, \mathcal{Z})$, respectively, which we assume are such that $H(X|Z) > H(X|Y)$ to support key generation. Let $\mathbf{Y}$ and $\mathbf{Z}$ represent Bob's and Willie's observations, respectively, corresponding to Alice's sequence $\mathbf{X}$. Let $M \in \mathcal{M}$ represent the message generated by Alice from $\mathbf{X}$, which is sent to Bob through a public channel. Let $P_{M\mathbf{XYZ}}$ denote the joint distribution induced by the coding scheme and q_M denote the uniform distribution on $\mathcal{M}$. The objective of covert reconciliation is to generate the same sequence at Alice's and Bob's terminals while avoiding detection by the warden Willie. We say that a covert forward-reconciliation throughput R is achievable with covertness $\delta > 0$ if there exists a sequence of reconciliation protocols

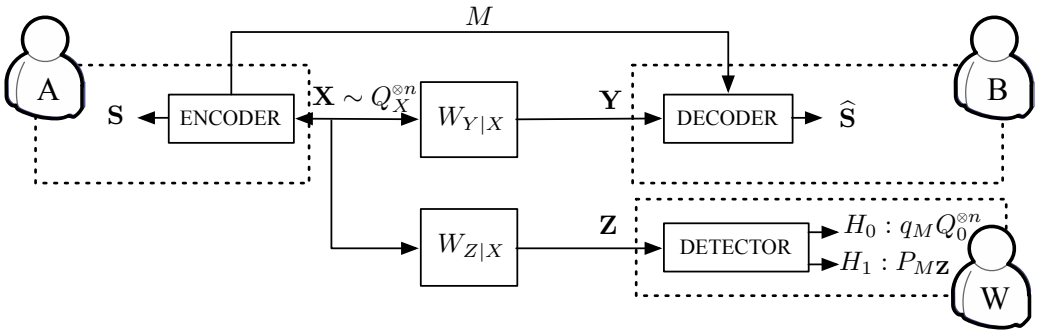

Figure 6.1: Channel model for covert forward-reconciliation.

$\{\mathcal{R}_n\}_{n\geqslant 1}$ such that

$$\lim_{n\to\infty} \mathbb{P}\Big(\mathbf{S} \neq \widehat{\mathbf{S}}|\mathcal{R}_n\Big) = 0, \tag{6.1}$$

$$\lim_{n\to\infty} \frac{H(\mathbf{S}|\mathcal{R}_n) - H(M|\mathcal{R}_n)}{\sqrt{n\delta}} \geqslant R, \tag{6.2}$$

$$\lim_{n\to\infty} \mathbb{D}(P_{M\mathbf{Z}} \| q_M Q_0^{\otimes n}) \leqslant \delta. \tag{6.3}$$

The supremum of all achievable throughputs is called the covert forward-reconciliation capacity. We refer the reader to [57] for a full discussion of the framework for covert key generation. The main benefit of focusing on reconciliation capacity is to delay secrecy considerations at a later stage of the protocol and analysis.

6.2 MLC-PPM for covert forward reconciliation

We now describe the construction of codes for covert forward-reconciliation using MLC and PPM, which is illustrated in Figure 6.1 and is similar to the setup in Chapter 3 for covert communication. Alice generates q binary sequences $\{\mathbf{X}_i\}_{i\in[\![1,q]\!]}$, each of length ℓ, according to a uniform distribution. We denote the j-th component of $\mathbf{X}_i$ by $X_{i,j}$. The q sequences together can be thought of as a two-dimensional array $\mathbf{X}$ with i-th row $\mathbf{X}_i = X_{i,1:\ell}$ and j-th column $X_{1:q,j}$. Each row is fed to an encoder and the PPM mapper as shown in Figure 3.1. The encoder corresponding to the sequence $\mathbf{X}_i$ generates binary sequences $\mathbf{M}_i$ and $\mathbf{K}_i$ from

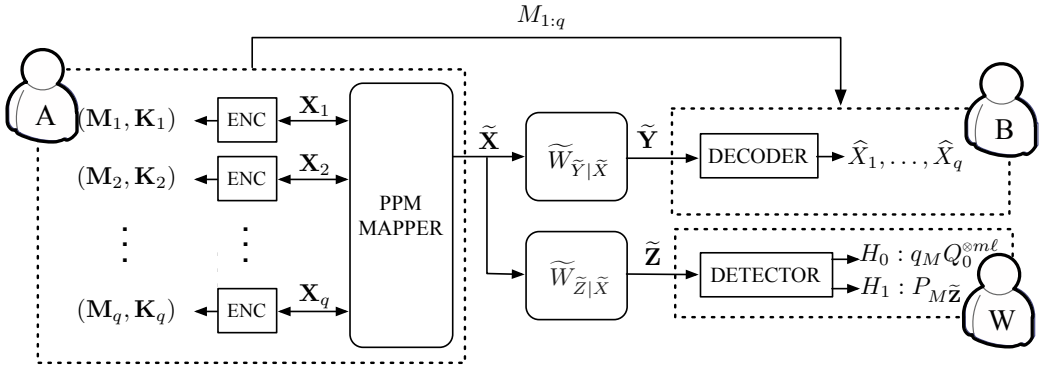

Figure 6.1: MLC-PPM setup for covert forward-reconciliation.

$\mathbf{X}_i$. Here, $\mathbf{M}_i$ is the message that Alice has to send to Bob to reconstruct the reconciliation sequence and $\mathbf{K}_i$ is the secret key that can be extracted on both Alice's and Bob's end. The PPM mapper of order q maps $X_{1:q,i}$ to $\widetilde{X}_i = \widetilde{x}(X_{1:q,i})$ to get the PPM sequence $\widetilde{\mathbf{X}}$ of length ℓ, which is observed by Bob and Willie at the output of their respective channels. Note that the j-th component of $\widetilde{\mathbf{X}}$ represented by $\widetilde{X}_j$ is a vector of length $m = 2^q$. We denote the output of the super channels corresponding to $\widetilde{\mathbf{X}}$ by $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$. Similar to $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{Y}}$ and $\widetilde{\mathbf{Z}}$ are sequences of length ℓ with each component of length m corresponding to one use of the super channel. The j-th component of $\widetilde{\mathbf{Y}}$ (similarly $\widetilde{\mathbf{Z}}$) is denoted by $\widetilde{Y}_j = (\widetilde{Y}_{j,1}, \ldots, \widetilde{Y}_{j,m})$.

We propose a block encoding scheme in which the keys extracted in one block is used in the next block to achieve covertness requirement for the public message. We denote the random variables associated with j-th block with a superscript as shown in the functional dependence graph for the scheme in Figure 6.1.

6.3 High-level analysis of MLC-PPM for covert forward reconciliation

We now provide a high-level analysis of the block coding scheme. Let $Q_{\widetilde{\mathbf{X}}\widetilde{\mathbf{Y}}\widetilde{\mathbf{Z}}}$ represent the distribution induced by a uniform distribution on the PPM symbols at the input of the channel. In the covert reconciliation scheme, as illustrated in Figure 3.1 for one block, Alice's encoders derive messages $\mathbf{M}_{1:q}$ and keys $\mathbf{K}_{1:q}$. We provide the details of an encod-

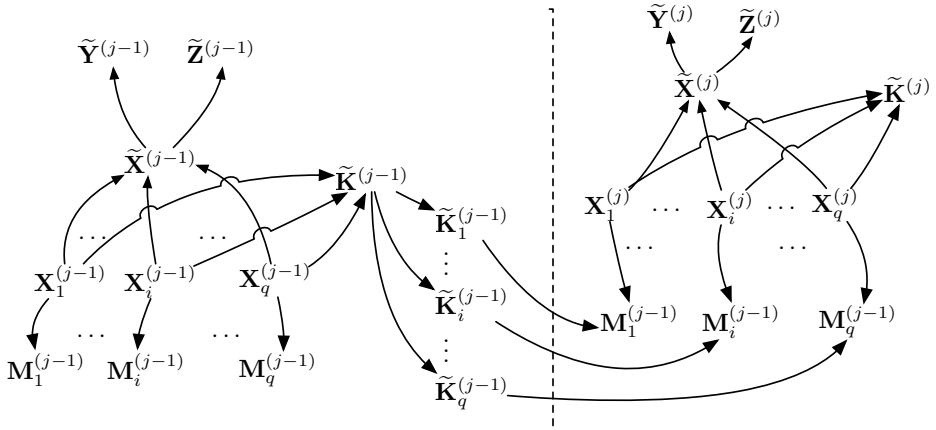

Figure 6.1: Functional dependence graph of the proposed block encoding scheme.

ing algorithm based on polar codes in the next section. The joint distribution of message $\mathbf{M}_{1:q}^{(1:B)}$ and Willie's output $\widetilde{\mathbf{Z}}^{(1:B)}$ induced by coding over B blocks is given by

$$P_{\mathbf{M}_{1:q}^{(1:B)}\widetilde{\mathbf{Z}}^{(1:B)}} = Q_{\widetilde{\mathbf{Z}}^{(1:B)}} \prod_{j=1}^{B} P_{\mathbf{M}_{1:q}^{(i)}|\mathbf{M}_{1:q}^{(i+1:B)}\widetilde{\mathbf{Z}}^{(i:B)}} \tag{6.4}$$

$$= \prod_{j=1}^{B} P_{\mathbf{M}_{1:q}^{(i)}\widetilde{\mathbf{Z}}^{(i)}|\mathbf{M}_{1:q}^{(i+1:B)}\widetilde{\mathbf{Z}}^{(i+1:B)}}. \tag{6.5}$$

We want to show that the relative entropy between the above distribution and the desired distribution $q_{\mathbf{M}_{1:q}^{(1:B)}} Q_0^{\otimes Bm\ell}$ can vanish. We have

$$\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(1:B)}\widetilde{\mathbf{Z}}^{(1:B)}} \| q_{\mathbf{M}_{1:q}^{(1:B)}} Q_0^{\otimes Bm\ell}\Big)$$

$$= \mathbb{E}_{P_{\mathbf{M}_{1:q}^{(1:B)}\widetilde{\mathbf{Z}}^{(1:B)}}}\left[\sum_{j=1}^{B}\left(\log\frac{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}}}{q_{\mathbf{M}_{1:q}^{(j)}} Q_0^{\otimes m\ell}} + \log\frac{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}|\mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)}}}{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}}}\right)\right] \tag{6.6}$$

$$= \sum_{j=1}^{B}\left(\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}} \| q_{\mathbf{M}_{1:q}^{(j)}} Q_0^{\otimes m\ell}\Big) + I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)})\right). \tag{6.7}$$

Moreover,

$$\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}} \| q_{\mathbf{M}_{1:q}^{(j)}} Q_0^{\otimes m\ell}\Big) = \mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}} \| q_{\mathbf{M}_{1:q}^{(j)}} Q_{\widetilde{\mathbf{Z}}^{(j)}}\Big) + \mathbb{D}\big(Q_{\widetilde{\mathbf{Z}}^{(j)}} \| Q_0^{\otimes m\ell}\big), \tag{6.8}$$

and

$$I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)}) \leqslant I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)}\widetilde{\mathbf{K}}_{1:q}^{(j)}) \tag{6.9}$$

$$= I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{K}}_{1:q}^{(j)}) + I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)}|\widetilde{\mathbf{K}}_{1:q}^{(j)}) \tag{6.10}$$

$$\overset{(a)}{=} I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{K}}_{1:q}^{(j)}), \tag{6.11}$$

where (a) follows because $\mathbf{M}_{1:q}^{(j+1:B)}\widetilde{\mathbf{Z}}^{(j+1:B)}$ depends on $\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}$ only through $\widetilde{\mathbf{K}}_{1:q}^{(j)}$ as shown in the functional dependence graph in Fig. 6.1.

Note that $Q_{\widetilde{\mathbf{Z}}^{(j)}}$ is the distribution induced at the output of Willie's super channel by an input uniformly distributed over PPM symbols. From [44], we know that

$$\mathbb{D}\big(Q_{\widetilde{\mathbf{Z}}^{(j)}} \| Q_0^{\otimes m\ell}\big) \leqslant \frac{\ell}{m}\left(\frac{\chi_2(Q_1||Q_0)}{2} + \mathcal{O}(\frac{1}{m})\right), \tag{6.12}$$

where

$$\chi_2(Q_1||Q_0) \triangleq \sum_{\mathcal{Z}} \frac{(Q_1(z) - Q_0(z))^2}{Q_0(z)} dz,$$

so that this term vanishes for an appropriate choice of ℓ and m.

In the next section, we will analyze the other two terms for a specific coding scheme based on polar codes. Most importantly, we now show that we can achieve the requirements of covert forward-reconciliation by coding for each level independently. As noted in Section 4.1, given that the higher levels $X_{i+1:q}$ are already decoded correctly, $\widetilde{Y}_{\mathcal{A}^q(X_{i+1:q})}$ is a sufficient statistic for decoding lower levels. Hence, we can represent the channel corresponding to each level by an equivalent channel when we use MSD at the receiver. The equivalent channel corresponding to i-th level for Bob's channel is given by

$$W_{\widetilde{Y}_{\mathcal{A}^q(X_{i+1:q})}|X_i}(y_1, \ldots, y_{2^i}|x_i) = \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1, \ldots, y_{2^i}) \frac{P_1(y_k)}{P_0(y_k)}. \tag{6.13}$$

To bound the terms in (6.8) and (6.11), we drop the superscript and analyze the terms $\mathbb{D}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}} \| q_{\mathbf{M}_{1:q}} Q_{\widetilde{\mathbf{Z}}}\Big)$ and $I(\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}; \widetilde{\mathbf{K}}_{1:q})$ for a single block. First, we analyze the term $\mathbb{D}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}} \| q_{\mathbf{M}_{1:q}} Q_{\widetilde{\mathbf{Z}}}\Big)$ by bounding $\mathbb{V}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}, q_{\mathbf{M}_{1:q}} Q_{\widetilde{\mathbf{Z}}}\Big)$. We have

$$P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}(\mathbf{m}_{1:q}, \widetilde{\mathbf{z}}) = \sum_{\mathbf{x}_{1:q}} P_{\mathbf{X}_{1:q}\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}(\mathbf{x}_{1:q}, \mathbf{m}_{1:q}, \widetilde{\mathbf{z}}) \tag{6.14}$$

$$= \sum_{\mathbf{x}_{1:q}} Q_{\mathbf{X}_{1:q}\widetilde{\mathbf{Z}}}(\mathbf{x}_{1:q}, \widetilde{\mathbf{z}}) \prod_{i=1}^{q} P_{\mathbf{M}_i|\mathbf{X}_i}(\mathbf{m}_i|\mathbf{x}_i). \tag{6.15}$$

For every $i \in [\![1, q]\!]$, define

$$P^{(i)}_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}(\mathbf{m}_{1:q}, \widetilde{\mathbf{z}}) \triangleq \sum_{\mathbf{x}_{1:q}} Q_{\mathbf{X}_{1:q}\widetilde{\mathbf{Z}}}(\mathbf{x}_{1:q}, \widetilde{\mathbf{z}}) \left(\prod_{j=1}^{i} q_{\mathbf{M}_j}\right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j|\mathbf{X}_j}(\mathbf{m}_j|\mathbf{x}_j)\right) \tag{6.16}$$

$$= \sum_{\mathbf{x}_{1:q}} \left(\prod_{j=1}^{i} q_{\mathbf{M}_j}(\mathbf{m}_j) Q_{\mathbf{X}_j}(\mathbf{x}_j)\right) \left(\prod_{j=i+1}^{q} P_{\mathbf{X}_j\mathbf{M}_j}(\mathbf{x}_j, \mathbf{m}_j)\right) \widetilde{W}^{\otimes \ell}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}). \tag{6.17}$$

The following lemma shows that we can achieve covertness condition by using a code that achieve covertness for the equivalent channel corresponding to each level.

Lemma 16. *For every $i \in [\![1, q]\!]$, consider the channel*

$$W_{\widetilde{Z}_{\mathcal{A}^q(X_{i+1:q})}|X_i}(z_1, \ldots, z_{2^i}|x_i) = \frac{1}{2^{i-1}} \sum_{k \in \mathcal{A}^i(x_i)} Q_0^{\otimes 2^i}(z_1, \ldots, z_{2^i}) \frac{Q_1(z_k)}{Q_0(z_k)}. \tag{6.18}$$

Let $P_{\mathbf{M}_i\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}$ denote the joint distribution induced by a code over this channel, and let $q_{\mathbf{M}_i} Q_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}$ denote the distribution in which the messages are uniformly distributed and independent of the output of the channel. If $\mathbb{V}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}, q_{\mathbf{M}_i} Q_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}\Big) \leqslant \delta_i$, then the same coding scheme ensures that $\mathbb{V}\Big(P^{(i-1)}_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}, P^{(i)}_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}\Big) \leqslant \delta_j$ irrespective of the coding scheme used for the higher levels. Moreover, for the output of the super channel $\widetilde{W}_{\widetilde{Z}|X_{1:q}}$, these codes together ensure $\mathbb{V}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}, q_{\mathbf{M}_{1:q}} Q_{\widetilde{\mathbf{Z}}}\Big) \leqslant \sum_{i=1}^{q} \delta_i$.

Proof. We have

$$\mathbb{V}\Big(P^{(i-1)}_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}, P^{(i)}_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}\Big)$$
$$= \sum_{\mathbf{m}_{1:q}} \sum_{\widetilde{\mathbf{z}}} \Bigg| \sum_{\mathbf{x}_{1:q}} \left(\prod_{j=1}^{i-1} q_{\mathbf{M}_j}(\mathbf{m}_j) Q_{\mathbf{X}_j}(\mathbf{x}_j)\right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j}(\mathbf{m}_j) P_{\mathbf{X}_j|\mathbf{M}_j}(\mathbf{x}_j|\mathbf{m}_j)\right)$$
$$\Big(q_{\mathbf{M}_i}(\mathbf{m}_i) Q_{\mathbf{X}_i(\mathbf{x_i})} - P_{\mathbf{M}_i\mathbf{X}_i}(\mathbf{m}_i, \mathbf{x}_i)\Big) \widetilde{W}^{\otimes \ell}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}) \Bigg| \tag{6.19}$$

$$\leqslant \sum_{\mathbf{m}_{1:q}} \sum_{\mathbf{x}_{i+1:q}} \left(\prod_{j=1}^{i-1} q_{\mathbf{M}_j}(m_j) \right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j}(\mathbf{m}_j) P_{\mathbf{X}_j|\mathbf{m}_j}(\mathbf{x}_j|\mathbf{m}_j) \right) \sum_{\widetilde{\mathbf{z}}} \Bigg| \sum_{\mathbf{x}_{1:i}} \left(\prod_{j=1}^{i-1} Q_{\mathbf{X}_j}(\mathbf{x}_j) \right) \Big(q_{\mathbf{M}_i}(\mathbf{m}_i) Q_{\mathbf{X}_i(\mathbf{x_i})} - P_{\mathbf{M}_i\mathbf{X}_i}(\mathbf{m}_i, \mathbf{x}_i) \Big) \widetilde{W}^{\otimes \ell}(\widetilde{\mathbf{z}}|\mathbf{x}_{1:q}) \Bigg| \tag{6.20}$$

$$= \sum_{\mathbf{m}_{1:q\setminus i}} \sum_{\mathbf{x}_{i+1:q}} \left(\prod_{j=1}^{i-1} q_{\mathbf{M}_j}(m_j) \right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j}(\mathbf{m}_j) P_{\mathbf{X}_j|\mathbf{m}_j}(\mathbf{x}_j|\mathbf{m}_j) \right) \sum_{\widetilde{\mathbf{z}} \setminus \widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}} \prod_{i'=1}^{\ell} \prod_{k' \in \mathcal{A}^q(x_{i+1:q}, i')^c} Q_0(\widetilde{z}_{i',k'}) \sum_{\mathbf{m}_i} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}} \Bigg| \sum_{\mathbf{x}_{1:i}} \left(\frac{1}{2^\ell} \right)^{i-1} \left(\frac{1}{|\mathbf{M}_i|} Q_{\mathbf{X}_i(\mathbf{x_i})} - P_{\mathbf{M}_i\mathbf{X}_i}(\mathbf{m}_i, \mathbf{x}_i) \right) \prod_{i''=1}^{\ell} Q_0^{\otimes 2^i}(\widetilde{z}_{i'', \mathcal{A}^q(x_{i+1:q,i''})}) \frac{Q_1(\widetilde{z}_{i, \mathcal{A}^q(x_{1:q,i''})})}{Q_0(\widetilde{z}_{i, \mathcal{A}^q(x_{1:q,i''})})} \Bigg| \tag{6.21}$$

$$= \sum_{\mathbf{m}_{1:q\setminus i}} \sum_{\mathbf{x}_{i+1:q}} \left(\prod_{j=1}^{i-1} q_{\mathbf{M}_j}(m_j) \right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j}(\mathbf{m}_j) P_{\mathbf{X}_j|\mathbf{m}_j}(\mathbf{x}_j|\mathbf{m}_j) \right) \sum_{\widetilde{\mathbf{z}} \setminus \widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}} \prod_{i'=1}^{\ell} \prod_{k' \in \mathcal{A}^q(x_{i+1:q}, i')^c} Q_0(\widetilde{z}_{i',k'}) \sum_{\mathbf{m}_i} \sum_{\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}} \Bigg| \frac{1}{|\mathbf{M}_i|} \sum_{\mathbf{x}_i} Q_{\mathbf{X}_i(\mathbf{x_i})} W^{\otimes \ell}_{\widetilde{Z}_{\mathcal{A}^q(X_{i+1:q})}|X_i}(\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}|\mathbf{x}_i) - \sum_{\mathbf{x}_i} P_{\mathbf{M}_i|\mathbf{X}_i}(\mathbf{m}_i|\mathbf{x}_i) Q_{\mathbf{X}_i(\mathbf{x_i})} W^{\otimes \ell}_{\widetilde{Z}_{\mathcal{A}^q(X_{i+1:q})}|X_i}(\widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}|\mathbf{x}_i) \Bigg| \tag{6.22}$$

$$= \sum_{\mathbf{m}_{1:q\setminus i}} \sum_{\mathbf{x}_{i+1:q}} \left(\prod_{j=1}^{i-1} q_{\mathbf{M}_j}(m_j) \right) \left(\prod_{j=i+1}^{q} P_{\mathbf{M}_j}(\mathbf{m}_j) P_{\mathbf{X}_j|\mathbf{m}_j}(\mathbf{x}_j|\mathbf{m}_j) \right) \sum_{\widetilde{\mathbf{z}} \setminus \widetilde{\mathbf{z}}_{\mathcal{A}^q(\mathbf{x}_{i+1:q})}} \prod_{i'=1}^{\ell} \prod_{k' \in \mathcal{A}^q(x_{i+1:q}, i')^c} Q_0(\widetilde{z}_{i',k'}) \mathbb{V}\Big(P_{\mathbf{M}_i \widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}, q_{\mathbf{M}_i} Q_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}} \Big) \tag{6.23}$$

$$= \mathbb{V}\Big(P_{\mathbf{M}_i \widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}, q_{\mathbf{M}_i} Q_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}} \Big), \tag{6.24}$$

where (6.24) follows because $\mathbb{V}\Big(P_{\mathbf{M}_i \widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}, q_{\mathbf{M}_i} Q_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}} \Big)$ is same for all realizations of $\mathbf{X}_{i+1:q}$ and $\mathbf{M}_{1:q\setminus i}$. □

We can also show that the term on the right hand side of (6.11) can be ensured to be

small by using codes that achieve a similar condition for the equivalent channel, which is summarized in the following lemma.

Lemma 17. *Suppose we have a code that ensure* $\mathbb{D}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{K}}_i\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}P_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}\Big) \leqslant \delta_i$ *for the equivalent channel in* (6.18) *for each* $i \in [\![1, q]\!]$*, then the overall coding scheme ensures* $I(\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}; \widetilde{\mathbf{K}}_{1:q}) \leqslant \sum_{i=1}^{q} \delta_i$.

Proof.

$$
\begin{aligned}
I(\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}; \widetilde{\mathbf{K}}_{1:q}) &= \mathbb{D}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}\|P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}P_{\widetilde{\mathbf{K}}_{1:q}}\Big) && (6.25)\\
&= \mathbb{E}_{P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}}\left[\log\frac{P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}}{q_{\mathbf{M}_{1:q}}q_{\widetilde{\mathbf{K}}_{1:q}}P_{\widetilde{\mathbf{Z}}}} + \log\frac{q_{\mathbf{M}_{1:q}}q_{\widetilde{\mathbf{K}}_{1:q}}P_{\widetilde{\mathbf{Z}}}}{P_{\mathbf{M}_{1:q}\widetilde{\mathbf{Z}}}P_{\widetilde{\mathbf{K}}_{1:q}}}\right] && (6.26)\\
&= \mathbb{D}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}\|q_{\mathbf{M}_{1:q}}q_{\widetilde{\mathbf{K}}_{1:q}}P_{\widetilde{\mathbf{Z}}}\Big) + H(\widetilde{\mathbf{K}}_{1:q}) + H(\mathbf{M}_{1:q}|\widetilde{\mathbf{Z}})\\
&\qquad + \log q_{\mathbf{M}_{1:q}} + \log q_{\widetilde{\mathbf{K}}_{1:q}} && (6.27)\\
&\overset{(a)}{\leqslant} \mathbb{D}\Big(P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}\|q_{\mathbf{M}_{1:q}}q_{\widetilde{\mathbf{K}}_{1:q}}P_{\widetilde{\mathbf{Z}}}\Big) && (6.28)\\
&\overset{(b)}{=} \sum_{i=1}^{q}\mathbb{D}\Big(P_{\mathbf{M}_{i:q}\widetilde{\mathbf{K}}_{i:q}\widetilde{\mathbf{Z}}}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}P_{\mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}\widetilde{\mathbf{Z}}}\Big) && (6.29)\\
&= \sum_{i=1}^{q}\mathbb{D}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{K}}_i}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}\Big) + I(\mathbf{M}_i\widetilde{\mathbf{K}}_i; \mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}\widetilde{\mathbf{Z}}) && (6.30)\\
&\overset{(c)}{\leqslant} \sum_{i=1}^{q}\mathbb{D}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{K}}_i}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}\Big) + I(\mathbf{M}_i\widetilde{\mathbf{K}}_i; \mathbf{X}_{i+1:q}\widetilde{\mathbf{Z}}) && (6.31)\\
&\overset{(d)}{=} \sum_{i=1}^{q}\mathbb{D}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{K}}_i}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}\Big) + I(\mathbf{M}_i\widetilde{\mathbf{K}}_i; \widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}) && (6.32)\\
&= \sum_{i=1}^{q}\mathbb{D}\Big(P_{\mathbf{M}_i\widetilde{\mathbf{K}}_i\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}\|q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}P_{\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}}\Big) && (6.33)\\
&\leqslant \sum_{i=1}^{q}\delta_i, && (6.34)
\end{aligned}
$$

where (a) follows because uniform distribution is the one that maximizes the entropy and therefore, $H(\widetilde{\mathbf{K}}_{1:q}) + \log q_{\widetilde{\mathbf{K}}_{1:q}} \leqslant 0$ and $H(\mathbf{M}_{1:q}|\widetilde{\mathbf{Z}}) + \log q_{\mathbf{M}_{1:q}} \leqslant 0$, (b) follows from

expressing $\frac{P_{\mathbf{M}_{1:q}\widetilde{\mathbf{K}}_{1:q}\widetilde{\mathbf{Z}}}}{q_{\mathbf{M}_{1:q}}q_{\widetilde{\mathbf{K}}_{1:q}}P_{\widetilde{\mathbf{Z}}}} = \prod_{i=1}^{q}\frac{P_{\mathbf{M}_{i:q}\widetilde{\mathbf{K}}_{i:q}\widetilde{\mathbf{Z}}}}{q_{\mathbf{M}_i}q_{\widetilde{\mathbf{K}}_i}P_{\mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}\widetilde{\mathbf{Z}}}}$, (c) follows from noting that

$$I(\mathbf{M}_i\widetilde{\mathbf{K}}_i;\mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}\widetilde{\mathbf{Z}}) \leqslant I(\mathbf{M}_i\widetilde{\mathbf{K}}_i;\mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}\mathbf{X}_{i+1:q}\widetilde{\mathbf{Z}})$$

and $I(\mathbf{M}_i\widetilde{\mathbf{K}}_i;\mathbf{M}_{i+1:q}\widetilde{\mathbf{K}}_{i+1:q}|\mathbf{X}_{i+1:q}\widetilde{\mathbf{Z}}) = 0$, and (d) follows because given the realizations of $\mathbf{X}_{i+1:q}$, knowledge of $\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}$ is enough for the decoding of $\mathbf{M}_i$ and $\widetilde{\mathbf{K}}_i$. □

Similarly, the reliability condition is achieved by using a code that achieves reliability for the equivalent channel corresponding to each level.

Lemma 18. *Suppose we have an encoder ϕ_i that encodes a message $\mathbf{M}_i = \phi_i(\mathbf{X}_i)$ and a decoder that reconstructs $\mathbf{X}_i$ as $\widehat{\mathbf{X}}_i = \psi_i(\mathbf{M_i}, \widetilde{\mathbf{Y}})$ from the message $\mathbf{M}_i$ and the output of the channel*

$$W_{\widetilde{Y}_{\mathcal{A}^q(X_{i+1:q})}|X_i}(y_1,\ldots,y_{2^i}|x_i) = \frac{1}{2^{i-1}}\sum_{k\in\mathcal{A}^i(x_i)} P_0^{\otimes 2^i}(y_1,\ldots,y_{2^i})\frac{P_1(y_k)}{P_0(y_k)}, \tag{6.35}$$

such that $\mathbb{P}\Big(\widehat{\mathbf{X}}_i \neq \mathbf{X}_i\Big) \leqslant \epsilon_i$, then there exists an encoder-decoder pair for the MLC-PPM scheme such that $\mathbb{P}\Big(\widehat{\mathbf{X}}_{1:q} \neq \mathbf{X}_{1:q}\Big) \leqslant \sum_{i=1}^{q}\epsilon_i$.

The proof of this lemma is similar to the proof of Lemma 10. For understanding the difference in the proofs, we include it here.

Proof. Let $\mathcal{E}_i \triangleq \{\widehat{\mathbf{X}}_i \neq \mathbf{X}_i\}$. We can expand the probability of error as

$$\mathbb{P}\Big((\widehat{\mathbf{X}}_1,\ldots,\widehat{\mathbf{X}}_q) \neq (\mathbf{X}_1,\ldots,\mathbf{X}_q)\Big) = \sum_{i=1}^{q}\mathbb{P}\left(\mathcal{E}_i\bigcap\bigcap_{j=i+1}^{q}\mathcal{E}_j^c\right) \tag{6.36}$$

We expand each term in the summation as

$$\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right) = \mathbb{P}\left(\widehat{\mathbf{X}}_i \neq \mathbf{X}_i, \widehat{\mathbf{X}}_{i+1} = \mathbf{X}_{i+1}, \ldots, \widehat{\mathbf{X}}_q = \mathbf{X}_q\right) \tag{6.37}$$

$$= \sum_{\mathbf{x}_{i:q}} \mathbb{P}(\mathbf{X}_i = \mathbf{x}_i)\mathbb{P}\left(\widehat{\mathbf{X}}_{i+1} = \mathbf{X}_{i+1} = \mathbf{x}_{i+1}, \ldots, \widehat{\mathbf{X}}_q = \mathbf{X}_q = \mathbf{x}_q\right)$$

$$\mathbb{P}\left(\widehat{\mathbf{X}}_i = \mathbf{x}_i | \mathbf{X}_i = \mathbf{x}_i, \widehat{\mathbf{X}}_{i+1} = \mathbf{X}_{i+1} = \mathbf{x}_{i+1}, \ldots, \widehat{\mathbf{X}}_q = \mathbf{X}_q = \mathbf{x}_q\right) \tag{6.38}$$

We now analyze the conditional probability term in the above summation as follows.

$$\mathbb{P}\left(\widehat{\mathbf{X}}_i = \mathbf{x}_i | \mathbf{X}_i = \mathbf{x}_i, \widehat{\mathbf{X}}_{i+1} = \mathbf{X}_{i+1} = \mathbf{x}_{i+1}, \ldots, \widehat{\mathbf{X}}_q = \mathbf{X}_q = \mathbf{x}_q\right)$$

$$= \sum_{\mathbf{m}_i, \widetilde{\mathbf{y}}: \psi_i(\mathbf{m}_i, \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \mathbb{P}\left(\mathbf{M}_i = \mathbf{m}_i, \widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})} = \widetilde{\mathbf{y}} | \mathbf{X}_i = \mathbf{x}_i, \ldots, \mathbf{X}_q = \mathbf{x}_q\right) \tag{6.39}$$

$$= \sum_{\mathbf{m}_i, \widetilde{\mathbf{y}}: \psi_i(\mathbf{m}_i, \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \mathbb{P}(\mathbf{M}_i = \mathbf{m}_i | \mathbf{X}_i = \mathbf{x}_i)\mathbb{P}\left(\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})} = \widetilde{\mathbf{y}} | \mathbf{X}_i = \mathbf{x}_i, \ldots, \mathbf{X}_q = \mathbf{x}_q\right) \tag{6.40}$$

$$= \sum_{\widetilde{\mathbf{y}}: \psi_i(\phi_i(\mathbf{x}_i), \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \mathbb{P}\left(\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})} = \widetilde{\mathbf{y}} | \mathbf{X}_i = \mathbf{x}_i, \ldots, \mathbf{X}_q = \mathbf{x}_q\right) \tag{6.41}$$

$$= \sum_{\widetilde{\mathbf{y}}: \psi_i(\phi_i(\mathbf{x}_i), \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1})$$

$$\prod_{j=1}^{\ell} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j, \mathcal{A}^i(x_{1:i,j})})}{P_0(\widetilde{y}_{j, \mathcal{A}^i(x_{1:i,j})})} \tag{6.42}$$

$$= \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1})$$

$$\sum_{\widetilde{\mathbf{y}}: \psi_i(\phi_i(\mathbf{x}_i), \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \prod_{j=1}^{\ell} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j, \mathcal{A}^i(x_{1:i,j})})}{P_0(\widetilde{y}_{j, \mathcal{A}^i(x_{1:i,j})})} \tag{6.43}$$

$$= \sum_{\mathbf{x}_1} \cdots \sum_{\mathbf{x}_{i-1}} \mathbb{P}(\mathbf{X}_1 = \mathbf{x}_1, \ldots, \mathbf{X}_{i-1} = \mathbf{x}_{i-1})$$

$$\sum_{\widetilde{\mathbf{y}}: \psi_i(\phi_i(\mathbf{x}_i), \widetilde{\mathbf{y}}) \neq \mathbf{x}_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k \in \mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j,k})}{P_0(\widetilde{y}_{j,k})} \tag{6.44}$$

$$= \sum_{\widetilde{\mathbf{y}}:\psi_i(\phi_i(\mathbf{x}_i),\widetilde{\mathbf{y}})\neq\mathbf{x}_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j,k})}{P_0(\widetilde{y}_{j,k})}. \tag{6.45}$$

Similar to the analysis in the proof of Lemma 3, we have

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right) \\
&\quad \leqslant \sum_{\mathbf{x}_{i:q}} \mathbb{P}(\mathbf{X}_i = \mathbf{x}_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\phi_i(\mathbf{x}_i),\widetilde{\mathbf{y}})\neq\mathbf{x}_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j,k})}{P_0(\widetilde{y}_{j,k})}.
\end{aligned} \tag{6.46}$$

The probability of error for the reconstruction of source using the output of the equivalent channel as the side-information is given by

$$\mathbb{P}\Big(\widehat{\mathbf{X}}_i \neq \mathbf{X}_i\Big) = \sum_{\mathbf{x}_i} \mathbb{P}(\mathbf{X}_i = \mathbf{x}_i) \mathbb{P}\Big(\widehat{\mathbf{X}}_i \neq \mathbf{x}_i | \mathbf{X}_i = \mathbf{x}_i\Big) \tag{6.47}$$

$$= \sum_{\mathbf{x}_i} \mathbb{P}(\mathbf{X}_i = \mathbf{x}_i) \sum_{\widetilde{\mathbf{y}}:\psi_i(\phi_i(\mathbf{x}_i),\widetilde{\mathbf{y}})\neq\mathbf{x}_i} \left(\frac{1}{2^{i-1}}\right)^{\ell} \prod_{j=1}^{\ell} \sum_{k\in\mathcal{A}^i(x_{i,j})} P_0^{\otimes 2^i}(\widetilde{y}_j) \frac{P_1(\widetilde{y}_{j,k})}{P_0(\widetilde{y}_{j,k})} \tag{6.48}$$

Hence, a code that guarantees $\mathbb{P}\Big(\widehat{\mathbf{X}}_i \neq \mathbf{X}_i\Big) \leqslant \epsilon_i$ for the equivalent channel satisfies

$$\mathbb{P}\left(\mathcal{E}_i \bigcap \bigcap_{j=i+1}^{q} \mathcal{E}_j^c\right) \leqslant \mathbb{P}\Big(\widehat{\mathbf{X}}_i \neq \mathbf{X}_i\Big) \leqslant \epsilon_i. \tag{6.49}$$

Therefore,

$$\mathbb{P}\Big((\widehat{\mathbf{X}}_1, \ldots, \widehat{\mathbf{X}}_q) \neq (\mathbf{X}_1, \ldots, \mathbf{X}_q)\Big) = \sum_{i=1}^{q} \epsilon_i. \tag{6.50}$$

□

In the next section, we provide an instantiation of a coding scheme based on source polarization.

6.4 Polar code for covert forward reconciliation

We now describe a coding scheme for covert forward-reconciliation based on polar codes. Let the input to the i-th level be $\mathbf{X}_i$. Let the polar transform of $\mathbf{X}_i$ be $\mathbf{U}_i \triangleq G_\ell \mathbf{X}_i$, where G_ℓ is the source polarization transformation defined in [59]. For the equivalent channels $W_{\widetilde{Y}_{\mathcal{A}^q(X_{i+1:q})}|X_i}$ and $W_{\widetilde{Z}_{\mathcal{A}^q(X_{i+1:q})}|X_i}$, we define the following sets:

$$\mathcal{H}_{X_i|\widetilde{Y}} \triangleq \{k \in [\![1,\ell]\!] : H(U_{i,k}|U_{i,1:k-1}\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}) > \delta_\ell\} \tag{6.51}$$

$$\mathcal{V}_{X_i|\widetilde{Z}} \triangleq \{k \in [\![1,\ell]\!] : H(U_{i,k}|U_{i,1:k-1}\widetilde{\mathbf{Z}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}) > 1-\delta_\ell\}, \tag{6.52}$$

where $\delta_\ell \triangleq 2^{-\ell^\beta}$ for $0 < \beta < 1/2$.

From the definition of the equivalent channel, the memoryless nature of the channels, and the fact that the polarization transform is invertible, we have the Markov chain $U_{i,k} - (U_{i,1:k-1}, \widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}) - (\mathbf{U}_{i+1:q}, \widetilde{\mathbf{Y}} \backslash \widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})})$. Hence,

$$H(U_{i,k}|U_{i,1:k-1}\widetilde{\mathbf{Y}}_{\mathcal{A}^q(\mathbf{X}_{i+1:q})}) = H(U_{i,k}|U_{i,1:k-1}\mathbf{U}_{i+1:q}\widetilde{\mathbf{Y}}). \tag{6.53}$$

Algorithm 1: Alice's encoding

Require $\widetilde{\mathbf{K}}_{1:q}^{(0)}$
for *Block* $j = 1$ *to* k **do**
  for *Level* $i = 1$ *to* q **do**
    $\mathbf{U}_i^{(j)} = G_\ell \mathbf{X}_i^{(j)}$
    $\mathbf{K}_i^{(j)} = \mathbf{U}_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}]$
    $\mathbf{F}_i^{(j)} = \mathbf{U}_i^{(j)}[\mathcal{H}_{X_i|\widetilde{Y}} \bigcap \mathcal{V}_{X_i|\widetilde{Z}}]$
    $\widetilde{\mathbf{F}}_i^{(j)} = \mathbf{U}_i^{(j)}[\mathcal{H}_{X_i|\widetilde{Y}} \backslash \mathcal{V}_{X_i|\widetilde{Z}}]$
    $\mathbf{M}_i^{(j)} = [\mathbf{F}_i^{(j)}, \widetilde{\mathbf{F}}_i^{(j)} \oplus \widetilde{\mathbf{K}}_i^{(j-1)}]$
  end
  $\widetilde{\mathbf{K}}_{1:q}^{(j)} = \mathbf{K}_{1:q}^{(j)}[\mathcal{A}_{XYZ}]$
  Transmit $\mathbf{M}_{1:q}^{(j)}$
end

Alice's algorithm for extracting messages and keys from the generated sequence is

given in Algorithm 1. The encoding scheme requires a key $\widetilde{\mathbf{K}}_{1:q}^{(0)}$ for the first block to ensure covertness. For the subsequent blocks, Alice and Bob use a subset $\mathcal{A}_{XYZ}$ of the keys extracted in the previous block. Bob reconstructs Alice's sequence from his observations, the received messages, and the shared secret key as given in Algorithm 2. Note that the public message is split between $\mathbf{F}_i$ and $\widetilde{\mathbf{F}}_i$ to distinguish the bits that polarization makes nearly uniform from those it does not.

Algorithm 2: Bob's decoding

Require $\widetilde{\mathbf{K}}_{1:q}^{(0)}$
for *Block* $j = 1$ *to* k **do**
  for *Level* $i = q$ *to* 1 **do**
    Form $\mathbf{U}_i^{(j)}[\mathcal{H}_{X_i|\widetilde{Y}}]$ from $\mathbf{M}_i^{(j)}$ and $\widehat{\widetilde{\mathbf{K}}}_i^{(j-1)}$
    Compute $\widehat{\mathbf{U}}_i^{(j)}$ using the successive cancellation decoder [59]
    $\widehat{\mathbf{X}}_i^{(j)} = G_\ell \widehat{\mathbf{U}}_i^{(j)}$
    $\widehat{\mathbf{K}}_i^{(j)} = \widehat{\mathbf{U}}_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}]$
  end
  $\widehat{\widetilde{\mathbf{K}}}^{(j)} = \widehat{\mathbf{K}}_{1:q}^{(j)}[\mathcal{A}_{XYZ}]$
end

6.4.1 Existence of polarized sets

From our assumption, we have $H(\widetilde{X}|\widetilde{Z}) > H(\widetilde{X}|\widetilde{Y})$. By using [60, Lemma 1], we have

$$\lim_{\ell\to\infty} \frac{|\mathcal{H}_{X_i|\widetilde{Y}}|}{\ell} = H(X_i|X_{i+1:q}\widetilde{Y}), \qquad \lim \sum_{i=1}^{q} \frac{|\mathcal{H}_{X_i|\widetilde{Y}}|}{\ell} = H(X_{1:q}|\widetilde{Y}),$$
$$\lim_{\ell\to\infty} \frac{|\mathcal{V}_{X_i|\widetilde{Z}}|}{\ell} = H(X_i|X_{i+1:q}\widetilde{Z}), \qquad \lim \sum_{i=1}^{q} \frac{|\mathcal{V}_{X_i|\widetilde{Z}}|}{\ell} = H(X_{1:q}|\widetilde{Z}).$$

The length of the key extracted in the $(j-1)$-th block is

$$|\mathbf{K}_{1:q}^{(j-1)}| = \sum_{i=1}^{q} |\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}|. \tag{6.54}$$

The length of the key used in the j-th block is

$$|\mathbf{K}_{1:q}^{(j)}| = \sum_{i=1}^{q} |\mathcal{H}_{X_i|\widetilde{Y}} \backslash \mathcal{V}_{X_i|\widetilde{Z}}|. \tag{6.55}$$

Using the observation that $|\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}| - |\mathcal{H}_{X_i|\widetilde{Y}} \backslash \mathcal{V}_{X_i|\widetilde{Z}}| = |\mathcal{V}_{X_i|\widetilde{Z}}| - |\mathcal{H}_{X_i|\widetilde{Y}}|$, we have

$$|\mathbf{K}_{1:q}^{(j-1)}| - |\mathbf{K}_{1:q}^{(j)}| = \sum_{i=1}^{q} \left(|\mathcal{V}_{X_i|\widetilde{Z}}| - |\mathcal{H}_{X_i|\widetilde{Y}}| \right). \tag{6.56}$$

Hence, by an analysis similar to the one in the proof of Theorem 1, we get

$$\lim_{n\to\infty} \frac{|\mathbf{K}_{1:q}^{(j-1)}| - |\mathbf{K}_{1:q}^{(j)}|}{\ell} = H(X_{1:q}|\widetilde{Z}) - H(X_{1:q}|\widetilde{Y}). \tag{6.57}$$

Therefore, we can use a part of the extracted key $\mathbf{K}_{1:q}^{(j-1)}$ in the j-th block.

6.4.2 Reliability

Using the similar analysis similar to the one in [60], we obtain

$$\mathbb{P}\left(\widehat{\mathbf{X}}_{1:q}^{(1:B)} \neq \mathbf{X}_{1:q}^{(1:B)}\right) \leqslant \sum_{j=1}^{B} \mathbb{P}\left(\widehat{\mathbf{X}}_{1:q}^{(j)} \neq \mathbf{X}_{1:q}^{(j)}\right) \tag{6.58}$$

$$= \sum_{j=1}^{B} \mathbb{P}\left(\widehat{\mathbf{U}}_{1:q}^{(j)} \neq \mathbf{U}_{1:q}^{(j)}\right) \tag{6.59}$$

$$\leqslant \frac{B(B+1)}{2} \ell \delta_\ell. \tag{6.60}$$

6.4.3 Covertness

We prove the covertness of the polar coding scheme described in Algorithm 1 using the following lemmas.

Lemma 19. *For every block $j \in [\![1, B]\!]$, the keys $\widetilde{\mathbf{K}}_{1:q}^{(j)}$ and the public messages $\mathbf{M}_{1:q}^{(j)}$*

generated according to Algorithm 1 and the adversary's observations $\widetilde{\mathbf{Z}}^{(j)}$ *satisfies*

$$I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) \leqslant 2q\ell\delta_\ell. \tag{6.61}$$

Proof. We have

$$\begin{aligned} I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) &= I(\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{F}}_{1:q}^{(j)}\oplus\widetilde{\mathbf{K}}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) && (6.62)\\ &= I(\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) + I(\widetilde{\mathbf{F}}_{1:q}^{(j)}\oplus\widetilde{\mathbf{K}}_{1:q}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}|\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)}) && (6.63)\end{aligned}$$

We bound the first term as follows:

$$\begin{aligned} I(\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) &\leqslant I(\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)}) && (6.64)\\ &= H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)}) - H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)}|\mathbf{F}_{1:q}^{(j)},\widetilde{\mathbf{Z}}^{(j)}) && (6.65)\\ &\leqslant |\widetilde{\mathbf{K}}_{1:q}^{(j)}| + |\mathbf{K}_{1:q}^{(j)}| - H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)},\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) + H(\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) && (6.66)\\ &\leqslant |\widetilde{\mathbf{K}}_{1:q}^{(j)}| + |\mathbf{K}_{1:q}^{(j)}| + |\mathbf{F}_{1:q}^{(j)}| - H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)},\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) && (6.67)\\ &= \sum_{i=1}^{q}|\mathcal{V}_{X_i|\widetilde{Z}}| - H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)},\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}). && (6.68)\end{aligned}$$

We now analyze the term $H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)},\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)})$ as follows.

$$\begin{aligned} H(\widetilde{\mathbf{K}}_{1:q}^{(j)},\mathbf{K}_{1:q}^{(j)},\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) &= H(U_{1:q}^{(j)}[\mathcal{V}_{X_{1:q}|\widetilde{Z}}]|\widetilde{\mathbf{Z}}^{(j)}) && (6.69)\\ &= \sum_{i=1}^{q} H(U_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}}]|U_{i+1:q}^{(j)}[\mathcal{V}_{X_{i+1:q}|\widetilde{Z}}]\widetilde{\mathbf{Z}}^{(j)}) && (6.70)\\ &\geqslant \sum_{i=1}^{q} H(U_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}}]|U_{i+1:q}^{(j)}\widetilde{\mathbf{Z}}^{j}) && (6.71)\\ &= \sum_{i=1}^{q}\sum_{k\in\mathcal{V}_{X_i|\widetilde{Z}}} H(U_{i,k}^{(j)}|U_{i,1:k-1}^{(j)}U_{i+1:q}^{(j)}\widetilde{\mathbf{Z}}^{j}) && (6.72)\\ &\geqslant \sum_{i=1}^{q}\sum_{k\in\mathcal{V}_{X_i|\widetilde{Z}}} (1-\delta_\ell) && (6.73)\end{aligned}$$

$$= \sum_{i=1}^{q} |\mathcal{V}_{X_i|\widetilde{Z}}|(1-\delta_\ell). \tag{6.74}$$

Hence, we obtain

$$I(\mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}; \widetilde{\mathbf{K}}_{1:q}^{(j)}) \leqslant \sum_{i=1}^{q} |\mathcal{V}_{X_i|\widetilde{Z}}|\delta_\ell \leqslant q\ell\delta_\ell \tag{6.75}$$

We bound the second term in (6.63) as follows:

$$\begin{aligned}
I(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)}; \widetilde{\mathbf{K}}_{1:q}^{(j)} | \mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) &\leqslant I(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)}; \widetilde{\mathbf{K}}_{1:q}^{(j)}, \mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{F}}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) && (6.76)\\
&= H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)}) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)} | \widetilde{\mathbf{K}}_{1:q}^{(j)}, \mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{F}}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) && (6.77)\\
&= H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)}) - H(\widetilde{\mathbf{K}}_{1:q}^{(j)} | \widetilde{\mathbf{K}}_{1:q}^{(j)}, \mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{F}}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) && (6.78)\\
&= H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j)}) - H(\widetilde{\mathbf{K}}_{1:q}^{(j)}) && (6.79)\\
&\leqslant |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - H(\widetilde{\mathbf{K}}_{1:q}^{(j)}) && (6.80)\\
&\leqslant |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - H(\mathbf{K}_{1:q}^{(j)}, \widetilde{\mathbf{K}}_{1:q}^{(j)}) && (6.81)\\
&= |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - H(U_{1:q}^{(j)}[\mathcal{V}_{X_{1:q}|\widetilde{Z}} \backslash \mathcal{H}_{X_{1:q}|\widetilde{Y}}]) && (6.82)\\
&= |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| \\
&\quad - \sum_{i=1}^{q} H(U_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}] | U_{i+1:q}^{(j)}[\mathcal{V}_{X_{i+1:q}|\widetilde{Z}} \backslash \mathcal{H}_{X_{i+1:q}|\widetilde{Y}}]) && (6.83)\\
&= |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - \sum_{i=1}^{q} H(U_i^{(j)}[\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}]) && (6.84)\\
&\leqslant |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - \sum_{i=1}^{q} \sum_{k \in \mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}} H(U_{i,k}^{(j)} | U_{i,1:k-1}^{(j)}) && (6.85)\\
&\leqslant |\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}| - \sum_{i=1}^{q} |\mathcal{V}_{X_i|\widetilde{Z}} \backslash \mathcal{H}_{X_i|\widetilde{Y}}|(1-\delta_\ell) && (6.86)\\
&= \left(|\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}|\right) \delta_\ell \leqslant q\ell\delta_\ell && (6.87)
\end{aligned}$$

where (6.81) is because $|\mathbf{K}_{1:q}^{(j)}| \geqslant H(\mathbf{K}_{1:q}^{(j)} | \widetilde{\mathbf{K}}_{1:q}^{(j)})$.

From (6.63),(6.75), and (6.87), we obtain

$$I(\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)};\widetilde{\mathbf{K}}_{1:q}^{(j)}) \leqslant 2q\ell\delta_\ell. \tag{6.88}$$

□

Lemma 20. *For every block $j \in [\![1, B]\!]$, the joint distribution of the public messages $\mathbf{M}_{1:q}^{(j)}$ generated according to Algorithm 1 and the adversary's observations $\widetilde{\mathbf{Z}}^{(j)}$ satisfies*

$$\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}} \| q_{\mathbf{M}_{1:q}^{(j)}} Q_{\widetilde{\mathbf{Z}}^{(j)}}\Big) \leqslant q\ell\delta_\ell. \tag{6.89}$$

Proof.

$$\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}} \| q_{\mathbf{M}_{1:q}^{(j)}} Q_{\widetilde{\mathbf{Z}}^{(j)}}\Big) = \mathbb{E}_{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}}}\left[\log \frac{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}}}{q_{\mathbf{M}_{1:q}^{(j)}} Q_{\widetilde{\mathbf{Z}}^{(j)}}}\right] \tag{6.90}$$

$$= \mathbb{E}_{P_{\mathbf{M}_{1:q}^{(j)}\widetilde{\mathbf{Z}}^{(j)}}}\left[\log \frac{P_{\mathbf{M}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}}}{q_{\mathbf{M}_{1:q}^{(j)}}}\right] \tag{6.91}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - H(\mathbf{M}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) \tag{6.92}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - H(\mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}|\widetilde{\mathbf{Z}}^{(j)}) \tag{6.93}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - H(\mathbf{F}_{1:q}^{(j)}|\widetilde{\mathbf{Z}}^{(j)}) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}|\mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) \tag{6.94}$$

$$\leqslant |\mathbf{M}_{1:q}^{(j)}| - |\mathbf{F}_{1:q}^{(j)}|(1-\delta_\ell) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}|\mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) \tag{6.95}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - |\mathbf{F}_{1:q}^{(j)}|(1-\delta_\ell) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}) + I(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}; \mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) \tag{6.96}$$

$$\leqslant |\mathbf{M}_{1:q}^{(j)}| - |\mathbf{F}_{1:q}^{(j)}|(1-\delta_\ell) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}) + I(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}; \widetilde{\mathbf{F}}_{1:q}^{(j)}) \tag{6.97}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - |\mathbf{F}_{1:q}^{(j)}|(1-\delta_\ell) - H(\widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}|\widetilde{\mathbf{F}}_{1:q}^{(j)}) \tag{6.98}$$

$$= |\mathbf{M}_{1:q}^{(j)}| - |\mathbf{F}_{1:q}^{(j)}|(1-\delta_\ell) - H(\widetilde{\mathbf{K}}_{1:q}^{(j-1)}) \tag{6.99}$$

$$= |\mathbf{F}_{1:q}^{(j)}|\delta_\ell + |\widetilde{\mathbf{K}}_{1:q}^{(j-1)}| - H(\widetilde{\mathbf{K}}_{1:q}^{(j-1)}) \tag{6.100}$$

$$= |\mathbf{F}_{1:q}^{(j)}|\delta_\ell + \Big(|\mathbf{K}_{1:q}^{(j)}| + |\widetilde{\mathbf{K}}_{1:q}^{(j)}|\Big)\,\delta_\ell \tag{6.101}$$

$$\leqslant q\ell\delta_\ell \tag{6.102}$$

where (6.97) is because of the Markov chain $(\mathbf{F}_{1:q}^{(j)}, \widetilde{\mathbf{Z}}^{(j)}) - \widetilde{\mathbf{F}}_{1:q}^{(j)} - \widetilde{\mathbf{F}}_{1:q}^{(j)} \oplus \widetilde{\mathbf{K}}_{1:q}^{(j-1)}$. □

From (6.7),(6.8),(6.11),(6.12) and Lemma 19 and Lemma 20, we obtain

$$\mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(1:B)}\widetilde{\mathbf{Z}}^{(1:B)}} \| q_{\mathbf{M}_{1:q}^{(1:B)}} Q_0^{\otimes Bm\ell}\Big) \leqslant 3Bq\ell 2^{-\ell^\beta} + \frac{B\ell}{m}\left(\frac{\chi_2(Q_1||Q_0)}{2} + \mathcal{O}(\frac{1}{m})\right). \quad (6.103)$$

For ℓ and m large enough and B and m, the term $\frac{B\ell\chi_2(Q_1||Q_0)}{2m}$ dominates over the other terms and by choosing $\ell = \lceil \frac{2m\delta}{B\chi_2(Q_1||Q_0)} \rceil$, we have

$$\lim_{\ell\to\infty} \mathbb{D}\Big(P_{\mathbf{M}_{1:q}^{(1:B)}\widetilde{\mathbf{Z}}^{(1:B)}} \| q_{\mathbf{M}_{1:q}^{(1:B)}} Q_0^{\otimes Bm\ell}\Big) \leqslant \delta. \quad (6.104)$$

6.4.4 Reconciliation throughput

The reconciliation algorithm presented here agrees on the sequence $\mathbf{X}_{1:q}^{1:B}$ in $Bm\ell$ channel uses. Hence, the reconciliation throughput achieved by this scheme is given by

$$\frac{H(\mathbf{X}_{1:q}^{(1:B)}) - H(\mathbf{M}_{1:q}^{(1:B)})}{\sqrt{Bm\ell\delta}}. \quad (6.105)$$

Since $\mathbf{X}_{1:q}^{1:B}$ has i.i.d. uniform distribution, $H(\mathbf{X}_{1:q}^{1:B}) = B\ell H(X_{1:q}) = B\ell H(\widetilde{X})$.

$$H(\mathbf{M}_{1:q}^{(1:B)}) = B\sum_{i=1}^{q}\sum_{k\in\mathcal{H}_{X_i|\widetilde{Y}}} H(U_k|U_{1:k-1}\mathbf{U}_{i+1:q}\widetilde{\mathbf{Y}}) \quad (6.106)$$

$$\leqslant B\sum_{i=1}^{q} |\mathcal{H}_{X_i|\widetilde{Y}}| \quad (6.107)$$

Hence, the reconciliation throughput is

$$\frac{H(\mathbf{X}_{1:q}^{1:B}) - H(\mathbf{M}_{1:q}^{(1:B)})}{\sqrt{Bm\ell\delta}} \geqslant \frac{Bl H(\widetilde{X}) - B\sum_{i=1}^{q}|\mathcal{H}_{X_i|\widetilde{Y}}|}{\sqrt{Bm\ell\delta}} \quad (6.108)$$

$$\geqslant \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}}\left(H(\widetilde{X}) - \sum_{i=1}^{q}|\mathcal{H}_{X_i|\widetilde{Y}}|/\ell\right) \quad (6.109)$$

Therefore,

$$\lim_{\ell\to\infty} \frac{H(\mathbf{X}_{1:q}^{1:B}) - H(\mathbf{M}_{1:q}^{(1:B)})}{\sqrt{Bm\ell\delta}} \geqslant \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} \lim_{m\to\infty} I(\widetilde{X};\widetilde{Y}) \tag{6.110}$$

$$= \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} \mathbb{D}(P_1\|P_0), \tag{6.111}$$

which one can show is optimal.

The amount of keys used in B blocks by this coding scheme is given by

$$B|\mathbf{K}_{1:q}^{(j)}| = B\sum_{i=1}^{q} |\mathcal{H}_{X_i|\widetilde{Y}} \backslash \mathcal{V}_{X_i|\widetilde{Z}}|. \tag{6.112}$$

Hence,

$$\lim_{\ell\to\infty} \frac{B\sum_{i=1}^{q} |\mathcal{H}_{X_i|\widetilde{Y}} \backslash \mathcal{V}_{X_i|\widetilde{Z}}|}{\sqrt{Bm\ell\delta}} \leqslant \lim_{\ell\to\infty} \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} \sum_{i=1}^{q} |\mathcal{H}_{X_i|\widetilde{Y}}|/\ell \tag{6.113}$$

$$= \sqrt{\frac{2}{\chi_2(Q_1||Q_0)}} H(\widetilde{X}|\widetilde{Y}), \tag{6.114}$$

which can be shown to be optimal.

CHAPTER 7
CONCLUSION AND PERSPECTIVES

With increasing concerns over the privacy and security of communication technologies, covert communication proposes a way to hide sensitive communication. The last decade witnessed significant contributions in the information-theoretic analysis of covert communication, including the limits of covert communication for point-to-point and multiuser scenarios. However, the research on the development of codes for covert communication lagged behind the information-theoretic analysis. In this work, we tried to bridge this gap by showing the existence of practical coding schemes for covert communication.

One of the main hurdles in the development of covert communication is the requirement for a coding scheme with a low rate that arises because of the square root law. The square root law of covert communication states that we can transmit only $\mathcal{O}(\sqrt{n})$ bits in n channel uses by avoiding detection by an adversary. This suggests that we need to induce a heavily biased input distribution. One way to achieve this is by mapping a sequence of symbols with the required bias to a "super-symbol", and then coding over those super-symbols. In this regard, the optimality of PPM as a super-symbol was shown in [44]. However, a coding scheme over the PPM symbol requires non-binary codes with the alphabet size scaling with the length of the codeword, making it difficult to analyze. We overcame this with the use of MLC with PPM.

In the MLC-PPM scheme, the "super-channel" with non-binary input is transformed into multiple channels with binary input. Although the number of levels scales with the blocklength, we have shown that most of the capacity concentrates on the first few levels and the equivalent channel corresponding to each existing level remains the same when we increase the number of levels. Furthermore, we have shown that one can independently design codes for each level by constructing codes for the equivalent channels, which in

turn achieve overall reliability and channel resolvability under mild assumptions regarding some symmetry conditions on the decoder. In particular, one may use polar codes because the successive cancellation decoder for polar codes satisfies the symmetry requirement. Finally, since the first few levels concentrate most of the capacity, one may code a limited number of levels for reliability and only code the remaining levels for resolvability using a negligible number of key bits. We have also shown how to code for channel resolvability on all of those remaining levels at once using invertible extractors, further reducing the complexity of the design.

While the discussion has focused on binary-input DMCs, the proposed coding scheme achieves reliability and covertness over any DMC, but without reaching the covert capacity in general. Extending the coding scheme to achieve the covert capacity is certainly possible, for instance by allowing more than a single non-innocent symbol in the PPM scheme. This is demonstrated in the extension of the MLC-PPM scheme using bi-orthogonal PPM symbols for covert communication over additive white Gaussian channels. While the practical application of this scheme for covert communication may not be possible in the immediate future, our illustration of this scheme for covert forward-reconciliation could find a potential application for quantum key distribution.

REFERENCES

[1] C. Cachin, "An information-theoretic model for steganography," *Information and Computation*, 2004.

[2] *Spread spectrum communications handbook*. McGraw-Hill, 1994, p. 1228.

[3] B. A. Bash, D. Goeckel, and D. Towsley, "Limits of reliable communication with low probability of detection on awgn channels," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 9, pp. 1921–1930, Sep. 2013.

[4] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[5] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.

[6] M. Bloch, M. Hayashi, and A. Thangaraj, "Error-control coding for physical-layer secrecy," *Proceedings of the IEEE*, vol. 103, no. 10, pp. 1725–1746, Oct. 2015.

[7] E. Lehmann and J. P. Romano, *Testing Statistical Hypotheses*. Springer.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2006.

[9] P. H. Che, M. Bakshi, and S. Jaggi, "Reliable deniable communication: hiding messages in noise," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, pp. 2945–2949.

[10] M. Bloch, "A channel resolvability perspective on stealth communications," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 2535–2539.

[11] B. A. Bash, D. Goeckel, and D. Towsley, "LPD communication when the warden does not know when," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 606–610.

[12] J. Hou and G. Kramer, "Effective secrecy: reliability, confusion and stealth," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 601–605.

[13] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable, deniable and hidable communication," in *Proc. of Information Theory and Applications Workshop (ITA)*, Feb. 2014, pp. 1–10.

[14] S. Kadhe, S. Jaggi, M. Bakshi, and A. Sprintson, "Reliable, deniable, and hidable communication over multipath networks," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, USA, Jun. 2014, pp. 611–615.

[15] P. H. Che, S. Kadhe, M. Bakshi, C. Chan, S. Jaggi, and A. Sprintson, "Reliable, deniable and hidable communication: a quick survey," in *Proc. of IEEE Information Theory Workshop (ITW)*, Nov. 2014, pp. 227–231.

[16] L. Wang, G. W. Wornell, and L. Zheng, "Fundamental limits of communication with low probability of detection," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 3493–3503, Jun. 2016.

[17] M. R. Bloch, "Covert communication over noisy channels: a resolvability perspective," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2334–2354, May 2016.

[18] M. Tahmasbi and M. R. Bloch, "First- and second-order asymptotics in covert communication," *IEEE Transactions on Information Theory*, vol. 65, no. 4, pp. 2190–2212, Apr. 2019.

[19] K. S. K. Arumugam and M. R. Bloch, "Keyless covert communication over multiple-access channels," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, 2016, pp. 2229–2233.

[20] ——, "Covert communication over a k-user multiple-access channel," *IEEE Transactions on Information Theory*, vol. 65, no. 11, pp. 7020–7044, 2019.

[21] ——, "Covert communication over broadcast channels," in *Proc. of IEEE Information Theory Workshop (ITW)*, Kaohsiung, Taiwan, 2017, pp. 299–303.

[22] ——, "Embedding covert information in broadcast communications," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 10, pp. 2787–2801, 2019.

[23] K. S. Kumar Arumugam, M. R. Bloch, and L. Wang, "Covert communication over a physically degraded relay channel with non-colluding wardens," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, Jun. 2018, pp. 766–770.

[24] M. Tahmasbi and M. R. Bloch, "Covert secret key generation," in *Proc. of IEEE Conference on Communications and Network Security (CNS)*, Las Vegas, NV, USA, Oct. 2017, pp. 540–544.

[25] B. A. Bash, S. Guha, D. Goeckel, and D. Towsley, "Quantum noise limited optical communication with low probability of detection," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Istanbul, Turkey, 2013.

[26] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, *Nature Communications*, pp. 1–9,

[27] B. A. Bash, C. N. Gagatsos, A. Datta, and S. Guha, "Fundamental limits of quantum-secure covert optical sensing," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, 2017.

[28] J. M. Arrazola and V. Scarani, "Covert quantum communication," *Physical Review Letters*, vol. 117, no. 25, p. 250 503, Dec. 2016.

[29] L. Wang, "Optimal throughput for covert communication over a classical-quantum channel," in *Proc. of IEEE Information Theory Workshop*, Cambridge, UK, Sep. 2016, pp. 364–368.

[30] P. H. Che, M. Bakshi, C. Chan, and S. Jaggi, "Reliable deniable communication with channel uncertainty," in *Proc. of IEEE Information Theory Workshop (ITW)*, Nov. 2014, pp. 30–34.

[31] S. Lee and R. J. Baxley, "Achieving positive rate with undetectable communication over awgn and rayleigh channels," in *Proc. of IEEE International Conference on Communications (ICC)*, Jun. 2014, pp. 780–785.

[32] S. Lee, R. J. Baxley, J. B. McMahon, and R. Scott Frazier, "Achieving positive rate with undetectable communication over mimo rayleigh channels," in *Proc. of IEEE 8th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, Jun. 2014, pp. 257–260.

[33] S. Lee, R. J. Baxley, M. A. Weitnauer, and B. Walkenhorst, "Achieving undetectable communication," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 7, pp. 1195–1205, Oct. 2015.

[34] T. V. Sobers, B. A. Bash, D. Goeckel, S. Guha, and D. Towsley, "Covert communication with the help of an uninformed jammer achieves positive rate," in *2015 49th Asilomar Conference on Signals, Systems and Computers*, Nov. 2015, pp. 625–629.

[35] T. S. Han and S. Verdù, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, 1993.

[36] "Strong secrecy from channel resolvability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[37] M. R. Bloch, L. Luzzi, and J. Kliewer, "Strong coordination with polar codes," in *Proc. of 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Oct. 2012, pp. 565–571.

[38] R. A. Chou, M. R. Bloch, and J. Kliewer, "Empirical and strong coordination via soft covering with polar codes," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5087–5100, Jul. 2018.

[39] R. A. Chou and M. R. Bloch, "Polar coding for the broadcast channel with confidential messages: a random binning analogy," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2410–2429, May 2016.

[40] R. A. Chou, M. R. Bloch, and J. Kliewer, "Low-complexity channel resolvability codes for the symmetric multiple-access channel," in *Proc. of IEEE Information Theory Workshop (ITW)*, Hobart, Tasmania, Nov. 2014, pp. 466–470.

[41] Q. Zhang, M. Bakshi, and S. Jaggi, "Computationally efficient deniable communication," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2234–2238, ISBN: 978-1-5090-1806-2.

[42] G. Frèche, M. Bloch, and M. Barret, "Polar codes for covert communications over asynchronous discrete memoryless channels," *Entropy*, vol. 20, no. 3, p. 18, Dec. 2017.

[43] K. S. K. Arumugam and M. R. Bloch, "Keyless asynchronous covert communication," in *Proc. of IEEE Information Theory Workshop (ITW)*, Cambridge, UK, Sep. 2016, pp. 191–195.

[44] M. R. Bloch and S. Guha, "Optimal covert communications using pulse-position modulation," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2825–2829.

[45] T. K. Moon, *Error correction coding : mathematical methods and algorithms*. Wiley-Interscience, 2005, p. 755.

[46] E. Arikan, "Channel polarization: a method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[47] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[48] U. Wachsmann, R. Fischer, and J. Huber, "Multilevel codes: theoretical concepts and practical design rules," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1361–1391, Jul. 1999.

[49] I. Sason and S. Verdù, "f-divergence inequalities," *IEEE Transactions on Information Theory*, vol. 62, no. 11, pp. 5973–6006, Nov. 2016.

[50] M. Bellare and S. Tessaro, *Polynomial-time, semantically-secure encryption achieving the secrecy capacity*, Jan. 2012. arXiv: 1201.3160.

[51] L. Wang, *Private communication*.

[52] H. Zhang and T. A. Gulliver, "Biorthogonal pulse position modulation for time-hopping multiple access uwb communications," *IEEE Transactions on Wireless Communications*, vol. 4, no. 3, pp. 1154–1162, May 2005.

[53] H. Tyagi and A. Vardy, "Explicit capacity-achieving coding scheme for the gaussian wiretap channel," in *Proc. IEEE International Symposium on Information Theory (ISIT)*, Honolulu, HI, Jun. 2014, pp. 956–960.

[54] B. A. Bash, A. H. Gheorghe, M. Patel, J. L. Habif, D. Goeckel, D. Towsley, and S. Guha, "Quantum-secure covert communication on bosonic channels," *Nature Communications*, vol. 6, pp. –, 2015.

[55] A. Sheikholeslami, B. A. Bash, D. Towsley, D. Goeckel, and S. Guha, "Covert communication over classical-quantum channels," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain, Jul. 2016, pp. 2064–2068.

[56] J. M. Arrazola and R. Amiri, "Secret-key expansion from covert communication," *Phys. Rev. A*, vol. 97, p. 022 325, 2 Feb. 2018. arXiv: 1708.09103v1 [quant-ph].

[57] M. Tahmasbi and M. R. Bloch, "Framework for covert and secret key expansion over classical-quantum channels," *Phys. Rev. A*, vol. 99, p. 052 329, 5 May 2019.

[58] G. Brassard and L. Salvail, "Secret-key reconciliation by public discussion," in *Advances in Cryptology-Eurocrypt'93*, T. Helleseth, Ed., Springer-Verlag, 1993, pp. 411–423.

[59] E. Arikan, "Source polarization," in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, Austin, TX, USA, Jun. 2010, pp. 899–903.

[60] R. A. Chou, M. R. Bloch, and E. Abbe, "Polar coding for secret-key generation," *IEEE Transactions on Information Theory*, vol. 61, no. 11, pp. 6213–6237, Nov. 2015.